J Mol Evol (2004) 59:158–176
DOI: 10.1007/s00239-004-2610-2

JOURNAL OF MOLECULAR EVOLUTION

Phylogenetic Analysis of Proteins Associated in the Four Major Energy Metabolism Systems: Photosynthesis, Aerobic Respiration, Denitrification, and Sulfur Respiration

Takeshi Tomiki, Naruya Saitou

Division of Population Genetics, National Institute of Genetics, and Department of Genetics, School of Life Sciences, Graduate University for Advanced Studies, Mishima, Japan

Received: 22 September 2003 / Accepted: 28 November 2004

Abstract. The four electron transfer energy metabolism systems, photosynthesis, aerobic respiration, denitrification, and sulfur respiration, are thought to be evolutionarily related because of the similarity of electron transfer patterns and the existence of some homologous proteins. How these systems have evolved is elusive. We therefore conducted a comprehensive homology search using PSI-BLAST, and phylogenetic analyses were conducted for the three homologous groups (groups 1–3) based on multiple alignments of domains defined in the Pfam database. There are five electron transfer types important for catalytic reaction in group 1, and many proteins bind molybdenum. Deletions of two domains led to loss of the function of binding molybdenum and ferredoxin, and these deletions seem to be critical for the electron transfer pattern changes in group 1. Two types of electron transfer were found in group 2, and all its member proteins bind siroheme and ferredoxin. Insertion of the pyridine nucleotide disulfide oxidoreductase domain seemed to be the critical point for the electron transfer pattern change in this group. The proteins belonging to group 3 are all flavin enzymes, and they bind flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN). Types of electron transfer in this group are divergent, but there are two common characteristics. NAD(P)H works as an electron donor or acceptor, and FAD or FMN transfers electrons from/to NAD(P)H. Electron transfer functions might be added to these common characteristics by the addition of functional domains through the evolution of group 3 proteins. Based on the phylogenetic analyses in this study and previous studies, we inferred the phylogeny of the energy metabolism systems as follows: photosynthesis (and possibly aerobic respiration) and the sulfur/nitrogen assimilation system first diverged, then the sulfur/nitrogen dissimilation system was produced from the latter system.

Key words: Energy metabolism — Photosynthesis — Aerobic respiration — Denitrification — Sulfur respiration — Protein phylogeny — PSI-BLAST — Pfam — Tree superimposition

Introduction

Photosynthesis, aerobic respiration, denitrification, and sulfur respiration are energy generation metabolisms. Proton potential is generated through electron transfer, and ATP synthase produces ATP with the help of this potential in all four systems. Furthermore, the electron transporter in all four metabolism systems is quinone, and cytochrome *c* also works as an electron transporter except for sulfur respiration. Some homologous proteins are used in different metabolic systems. The cytochrome bc_1 complex, in aerobic respiration and denitrification,

Correspondence to: Naruya Saitou; *email:* nsaitou@genes.nig.ac.jp

and the cytochrome b_6f complex, in photosynthesis, are evolutionarily related (Berry et al. 2000; Hauska et al. 1988). Subunits I and II of cytochrome *c* oxidase in aerobic respiration were shown to be homologous with nitric oxide reductase subunit b and monomeric nitrous oxide reductase in denitrification (Mogi et al. 1998; Saraste and Castresana 1994; Zumft et al. 1992). Respiratory nitrate reductase, periplasmic nitrogen reductase, and polysulfide reductase are homologous between denitrification and sulfur respiration (Dias et al. 1999; Krafft et al. 1992; Moreno-Vivian et al. 1999).

It is of interest how these proteins and characteristics common to the four energy metabolism systems evolved. Evolutionary relationships of some proteins in the four metabolisms were already analyzed as we described above. However, an exhaustive phylogenetic analysis of proteins among the four metabolisms has not been done. We therefore conducted a molecular phylogenetic analysis of all the available amino acid sequences of those four energy metabolism systems using the homology search program PSI-BLAST (Altschul et al. 1997), because it can detect evolutionary relationships of highly diverged proteins (Park et al. 1998). It is possible that homologous proteins in the four energy metabolism systems were also used in other nitrogen and sulfur metabolisms (nitrate assimilation, nitrogen fixation and nitrification in nitrogen metabolism, and assimilation and dissimilation in sulfur metabolism) in addition to denitrification and sulfur respiration. Therefore, we also included amino acid sequences in these metabolisms for our analysis.

Materials and Methods

Amino Acid Sequence Retrieval

We retrieved amino acid sequences of proteins involved in photosynthesis, aerobic respiration, nitrogen metabolism including denitrification, nitrate assimilation, nitrogen fixation, and nitrification and sulfur metabolism including assimilation and dissimilation from SWISS-PROT (Boeckmann et al. 2003). We referred mainly to the pathway maps of oxidative phosphorylation, ATP synthesis, photosynthesis, nitrogen metabolism, and sulfur metabolism in KEGG (http://www.genome.ad.jp/kegg/; Kanehisa et al. 2002). V-type ATP synthase in eukaryotic cells is included in the map of ATP synthesis in KEGG. However, it is not involved in this metabolism (Wieczorek et al. 1999). Therefore, this enzyme was excluded from the present study. In total, we used 182 proteins retrieved by this procedure (see the supplementary information for the whole list of retrieved proteins).

We used those sequence data as queries and searched SWISS-PROT by using PSI-BLAST (Altschul et al. 1997) with SEG filter and 10 passes. Based on the PSI-BLAST search result, we defined homologous groups of query sequences by taking their union. The E value of 10^{-3} was used as cutoff to define homologous proteins. We found amino acid regions (domains) shared by all the query sequences in this way. Amino acid sequences retrieved through PSI-BLAST using these query sequences belonging to one homologous group are used for phylogenetic analysis. When the function of a protein was not clearly defined, and was listed only as "hypothetical protein" in SWISSPROT, these sequences were ignored in further analyses.

Phylogenetic Analysis

Because amino acid sequences which we analyzed are so divergent, we did not use the whole region of each protein for phylogenetic analysis so as to obtain reliable multiple alignments and phylogenetic trees. We defined conserved regions based on domains listed in the Pfam database (http://www.sanger.ac.uk/Software/Pfam/; Bateman et al. 2002). Pfam is a database of conserved amino acid sequence regions. Each conserved region is called a family. Multiple alignment of conserved region is constructed in each family and is available from the Pfam Web site. We used multiple alignments available in the Pfam database. When we encountered amino acid regions which are not defined by Pfam but were weakly homologous with a known Pfam domain, those amino acid sequences and corresponding Pfam domain sequences were multiply aligned using ClustalX (Thompson et al. 1997). Neighbor-joining trees (Saitou and Nei 1987) were constructed using MEGA2 (Kumar et al. 2001). Numbers of amino acid substitutions were computed for both Poisson correction and gamma distance with shape parameter 1 (default value).

Identification of Electron Transfer Patterns

In this study, we analyzed the relationship between amino acid sequences and catalytic features mainly focused on patterns of electron transfer, because electron transfer is essential for catalysis. Information on catalytic features, especially about patterns of electron transfer, were retrieved from both the SWISS-PROT database and bibliographies cited in this paper. We categorized proteins whose amino acid sequence data were used in this study into four categories as follows. Categories of amino acid sequence data are shown in Table 1. References cited in Table 1 are listed in supplementary information.

A. The amino acid sequence of the protein was experimentally determined, and its function is known. Moreover, electron transfer pattern in the protein has been revealed or predicted. References regarding its function are shown in Table 1.
B. The amino acid sequence of the protein was experimentally determined, and its function is known. References regarding its function are shown in Table 1.
C. The amino acid sequence of the protein was predicted only by translating genomic or cDNA nucleotide sequence, and its function was predicted by sequence similarity in SWISS-PROT.
D. The amino acid sequence of the protein was predicted only by translating genomic or cDNA nucleotide sequence, and there is no description of its function in SWISS-PROT.

Results

Homologous Groups

We found 24 homologous groups of proteins through the PSI-BLAST search. Because our objective is to discover the changes of metabolic pattern among the four energy metabolism systems and other nitrogen and sulfur metabolisms, we only used homologous

Table 1. Electron transfer patterns of proteins found to belong to three homologous groups

Enzyme	SWISS-PROT ID	Accession no.	Category	Reference(s)[a]
		Group 1		
NADH dehydrogenase	NUOG_ECOLI	P33602	C	
	NUOG_RICPR	Q9ZCF6	C	
	NUOG_SALTY	P33900	C	
	NUOG_STRCO	Q9XAR0	C	
	NUG2_RHIME	P56914	C	
	NUAM_ACACA	Q37373	C	
	NUAM_BOVIN	P15690	A	Ohnishi 1998
	NUAM_HUMAN	P28331	B	Chow et al. 1991
	NUAM_RECAM	O21241	C	
	NUAM_SOLTU	Q43644	B	Rasmusson et al. 1998
	NUAM_NEUCR	P24918	B	Preis et al. 1991
	NQO3_PARDE	P29915	B	Xu et al. 1992
	NQO3_THETH	Q56223	C	
Formate dehydrogenase	FDHA_METFO	P06131	B	Shuber et al. 1986
	FDHA_METJA	Q60314	C	
	FDHF_ECOLI	P07658	A	Boyington et al. 1997; Sauter et al. 1992
	FDOG_ECOLI	P32176	B	Abaibou et al. 1995
	FDNG_ECOLI	P24183	A	Jormakka et al. 2002
	FDXG_HAEIN	P46448	C	
Nitrate reductase	NAPA_DESDE	P81186	A	Dias et al. 1999
	NAPA_ALCEU	P39185	B	Siddiqui et al. 1993
	NAPA_ECOLI	P33937	B	Grove et al. 1996; Thomas et al. 1999
	NAPA_PARPN	Q56350	A	Berks et al. 1995; Breton et al. 1994; Butler et al. 1999; Moreno-Vivian et al. 1999
	NAPA_RHOSH	Q53176	B	Reyes et al. 1996
	NARB_SYNP7	P39458	B	Rubio et al. 1996
	NARB_SYNY3	P73448	B	Aichi et al. 2001
	NASA_KLEOX	Q06457	B	Lin et al. 1993
	NASC_BACSU	P42434	B	Ogawa et al. 1995
	NARG_BACSU	P42175	C	
	NARG_ECOLI	P09152	A	Blasco et al. 1989; Guigliarelli et al. 1992; Magalon et al. 1997, 1998; Moreno-Vivian et al. 1999
	NARZ_ECOLI	P19319	B	Blasco et al. 1990
Polysulfide reductase	PSRA_WOLSU	P31075	B	Krafft et al. 1992
Thiosulfate reductase	PHSA_SALTY	P37600	B	Heinzinger et al. 1995
Dimethyl sulfoxide reductase	DMSA_ECOLI	P18775	A	Cammack and Weiner 1990; Heffron et al. 2001; Rothery et al. 1999; Rothery and Weiner 1996; Trieber et al. 1996
	DMSA_HAEIN	P45004	C	
	DMSA_RHOCA	Q52675	A	Bray et al. 2000; McAlpine et al. 1998; Schneider et al. 1996; Stewart et al. 2000
	DMSA_RHOSH	Q57366	A	Li et al. 2000; Schindelin et al. 1996
Trimethylamine–*N*-oxide reductase	TORA_ECOLI	P33225	A	Gon et al. 2001
	TORA_SHEMA	O87948	B	Dos Santos et al. 1998
	TORZ_ECOLI	P46923	B	Gon et al. 2000
	TORZ_HAEIN	P44798	C	
Biotin sulfoxide reductase	BISC_ECOLI	P20099	B	Pierson and Campbell 1990
	BISC_RHOSH	P54934	A	Garton et al. 2000; Pollock and Barber 2001; Temple et al. 2000

Table 1. Continued

Enzyme	SWISS-PROT ID	Accession no.	Category	Reference(s)[a]
		Group 2		
Sulfite reductase [NADPH] hemoprotein	CYSI_ECOLI	P17846	A	Crane et al. 1995, 1997b, c
	CYSI_SALTY	P17845	B	Ostrowski et al. 1989b
	CYSI_THIRO	P52673	B	Bruhl et al. 1996
Sulfite reductase (ferredoxin)	SIR_SYNY3	P72854	C	
	SIR_SYNP7	P30008	B	Gisselmann et al. 1993
	-	Q42590	B	Bruhl et al. 1996
Sulfite reductase, assimilatory type	SIR_DESVH	Q05805	A	Tan and Cowan 1991
Ferredoxin–nitrite reductase	NIR_SYNP7	P39661	C	
	NIR_PHOLA	Q51879	B	Merchan et al. 1995
	NIR_SPIOL	P05314	B	Back et al. 1988, 1991
	NIR_BETVE	P38500	C	
Nitrite reductase [NAD(P)H]	NIR_NEUCR	P38681	A	Lafferty and Garrett 1974; Prodouz and Garrett 1981; Vega 1976
	NIR_EMENI	P22944	B	Johnstone et al. 1990
	NIRB_KLEOX	Q06458	B	Lin et al. 1993
	NASD_BACSU	P42435	B	Ogawa et al. 1995
	NIRB_ECOLI	P08201	B	Jayaraman et al. 1987; Peakman et al. 1990
Anaerobic sulfite reductase	ASRC_SALTY	P26476	B	Huang and Barrett 1991
Sulfite reductase, dissimilatory type	DSRA_ARCFU	Q59109	B	Dahl et al. 1993
	DSRA_CHRVI	O33998	B	Hipp et al. 1997
	DSRB_ARCFU	Q59110	B	Dahl et al. 1993
	DSVA_DESVH	P45574	B	Karkhoff-Schweizer et al. 1995; Pierik et al. 1992
	DSVB_DESVH	P45575	B	Karkhoff-Schweizer et al. 1995; Pierik et al. 1992
		Group 3		
CDP-6-deoxy-delta-3, 4-glucoseen reductase	ASCD_YERPE	P37911	A	Gassner et al. 1996; Johnson et al. 1996
Anaerobic sulfite reductase	ASRB_SALTY	P26475	B	Huang and Barrett 1991
Benzoate 1,2-dioxygenase	BENC_ACICA	P07771	A	Karlsson et al. 2002
Bifunctional P-450:NADPH-P450 reductase	CPXB_BACME	P14779	A	Haines et al. 2001; Hazzard et al. 1997; Li and Poulos 1997; Peterson et al. 1997; Ravichandran et al. 1993; Sevrioukova et al. 1996, 1999a, b; Yeom et al. 1995
Sulfite reductase [NADPH] flavoprotein	CYSJ_ECOLI	P38038	A	Eschenbrenner et al. 1995a, b; Gruez et al. 2000; Ostrowski et al. 1989a; Siegel et al. 1973
	MT10_YEAST	P39692	B	Hansen et al. 1994
Phenol hydroxylase P5 protein	DMPP_PSESP	P19734	A	Powlowski and Shingler 1990
Flavohemoprotein	FHP_CANNO	Q03331	B	Iwaasa et al. 1992
	FHP_YEAST	P39676	B	Zhu and Riggs 1992
	HMPA_ALCEU	P39662	B	Cramm et al. 1994
NADH oxidoreductase hcr	HCR_ECOLI	P75824	A	van den Berg et al. 2000
Fruit protein PKIWI502	K502_ACTCH	P43394	D	
Flavin reductase	LUXG_VIBHA	P16447	C	
NADH–cytochrome b5 reductase	MCR1_YEAST	P36060	B	Hahne et al. 1994
	NC5R_BOVIN	P07514	B	Ozols et al. 1984, 1985; Strittmatter et al. 1992
	NC5R_RAT	P20070	B	Pietrini et al. 1988; Zenno et al. 1990
	NC5R_YEAST	P38626	C	
Methane monooxygenase	MMOC_METCA	P22868	A	Lund et al. 1985

Table 1. Continued

Enzyme	SWISS-PROT ID	Accession no.	Category	Reference(s)[a]
NADPH–cytochrome P450 reductase	NCPR_HUMAN	P16435	B	Haniu et al. 1989
	NCPR_RAT	P00388	B	Porter and Kasper 1985; Porter et al. 1990
	NCPR_YEAST	P16603	B	Yabusaki et al. 1988
Naphthalene 1,2-dioxygenase system, ferredoxin–NAD(+) reductase component	NDOR_PSEPU	Q52126	B	Simon et al. 1993
Nitrate reductase	NIA_ASPNG	P36858	B	Unkles et al. 1992
	NIA_PHYIN	P39864	B	Pieterse et al. 1995
	NIA_PICAN	P49050	B	Avila et al. 1995
	NIA2_ARATH	P11035	A	Skipper et al. 2001
Nitric-oxide synthase	NOS_DROME	Q27571	B	Regulski and Tully 1995
	NOS1_RAT	P29476	B	Bredt et al. 1991
	NOS2_MOUSE	P29477	A	Aoyagi et al. 2001; Crane et al. 1997a, 1998, 1999; Ghosh et al. 1999; McMillan et al. 2000; Siddhanta et al. 1998
	NOS3_BOVIN	P29473	B	Lamas et al. 1992
	NOS3_HUMAN	P29474	B	Janssens et al. 1992
	NS2A_HUMAN	P35228	A	Fischmann et al. 1999; Li et al. 1999
Phthalate dioxygenase reductase	PDR_BURCE	P33164	A	Batie et al. 1991; Correll et al. 1992
	PHT2_PSEPU	Q05182	B	Nomura et al. 1992
Phenoxybenzoate dioxygenase	POBB_PSEPS	Q52186	B	Dehmel et al. 1995
	RFBI_SALTY	P26395	D	
Toluene-4-monooxygenase	TMOF_PSEME	Q03304	B	Yen and Karl 1992
Vanillate *O*-demethylase oxidoreductase	VANB_PSEPU	O54037	B	Venturi et al. 1998
Xylene monooxygenase	XYLA_PSEPU	P21394	A	Shaw and Harayama 1992
Toluate 1, 2-dioxygenase	XYLZ_PSEPU	P23101	B	Harayama et al. 1991
Putative dioxygenase	YEAX_ECOLI	P76254	C	
Ferredoxine–NADP reductase	FENR_ANASO	P21890		Arakaki et al. 1997; Martinez-Julvez et al. 1998, 2001; Mayoral et al. 2000; Morales et al. 2000; Serre et al. 1996
	FENR_SYNY3	Q55318	A	Arakaki et al. 1997; van Thor et al. 1999
	FENR_PEA	P10933	B	Newman and Gray 1988
	FENR_SPIOL	P00455		Jansen et al. 1988; Karplus et al. 1984
	FENR_AZOVI	Q44532	A	Arakaki et al. 1997; Isas and Burgess 1994; Sridhar Prasad et al. 1998
	FENR_BUCAP	Q9Z615	C	
	FENR_ECOLI	P28861	A	Arakaki et al. 1997; Bianchi et al. 1993; Ingelman et al. 1997

[a] References are given in the supplementary information.

groups including proteins belonging to more than one metabolism system in the further analyses. Homologous groups suitable for our objective were groups 1–8. Table 2 lists these eight groups. Homologous relationships of groups 5–8 have already been reported (Berry et al. 2000; Hauska et al. 1988; Mogi et al. 1998; Saraste and Castresana 1994). All the proteins in group 4 had a conserved domain defined as the Fer4 domain (4Fe–4S ferredoxin binding domain) by the Pfam database (Bateman et al. 2002). Unfortunately, this domain is very short (less than 50 amino acids), and we could not construct a reliable phylogenetic tree. We thus analyze groups 1–3 in the following.

Table 2. Eight homologous protein groups found in this study

ID	Name	Metabolism	No. of homologous proteins
	Group 1		
NUOG_ECOLI	NADH dehydrogenase chain G	O	
NARG_ECOLI	Respiratory nitrate reductase alpha chain	N_{DN}	
NARB_SYNY3	Assimilatory nitrate reductase (ferredoxin-dependent)	N_{DN}	
NASA_KLEOX	Assimilatory nitrate reductase (NADH-dependent)	N_{DN}	44
NAPA_ECOLI	Periplasmic nitrate reductase	N_{DN}	
PHSA_SALTY	Thiosulfate reductase PhsA subunit	S_{DS}	
PSRA_WOLSU	Polysulfide reductase chain A	S_R	
	Group 2		
NIR_SYNP7	Ferredoxin–nitrite reductase	N_{AS}	
NIRB_ECOLI	Nitrite reductase (NAD(P)H) large subunit	N_{AS}	
DSRA_ARCFU	Sulfite reductase (dissimilatory-type) alpha subunit	S_{DS}	
DSRB_ARCFU	Sulfite reductase (dissimilatory-type) beta subunit	S_{DS}	21
ASRC_SALTY	Anaerobic sulfite reductase subunit C	S_{AS}	
CYSI_ECOLI	Sulfite reductase (NADPH) beta component	S_{AS}	
SIR_SYNY3	Sulfite reductase (ferredoxin)	S_{AS}	
	Group 3		
FENR_SYNY3	Ferredoxin–NADP reductase	P	
NIA2_ARATH	Nitrate reductase [NADH]	N_{AS}	
NIA_BETVE	Nitrate reductase [NAD(P)H]	N_{AS}	122
NIA_NEUCR	Nitrate reductase [NADPH]	N_{AS}	
CYSJ_ECOLI	Sulfite reductase (NADPH) alpha component	S_{AS}	
ASRB_SALTY	Anaerobic sulfite reductase subunit B	S_{AS}	
	Group 4		
PSAC_SYNY3	Photosystem I subunit VII (PsaC)	P	
NUOI_ECOLI	NADH dehydrogenase chain I	O	
NARH_ECOLI	Respiratory nitrate reductase beta chain	N_{DN}	
PSRB_WOLSU	Polysulfide reductase chain B	S_R	185
PHSB_SALTY	Thiosulfate reductase PhsB subunit	S_{DS}	
DSRA_ARCFU	Sulfite reductase (dissimilatory-type) alpha subunit	S_{DS}	
DSRB_ARCFU	Sulfite reductase (dissimilatory-type) beta subunit	S_{DS}	
ASRC_SALTY	Anaerobic sulfite reductase subunit C	S_{AS}	
	Group 5		
CYB6_SYNY3	Cytochrome b6f complex cytochrome b6 subunit	P	956
CYB_YEAST	Cytochrome bcl complex cytochrome b subunit	O, N_{DN}	
	Group 6		
UCRI_SYNY3	Cytochrome b6f complex iron-sulfur subunit	P	37
UCRI_YEAST	Cytochrome bc1 complex iron-sulfur subunit	O, N_{DN}	
	Group 7		
COX2_YEAST	Cytochrome *c* oxidase polypeptide II	O	247
NOSZ_PSEST	Nitrous oxide reductase	N_{DN}	
	Group 8		
PETD_SYNY3	Cytochrome b6f complex subunit 4	P	937
CYB_YEAST	Cytochrome bc1 complex cytochrome *b* subunit	O, N_{DN}	

Note. P, photosynthesis; O, aerobic respiration; N_{DN}, denitrification; N_{AS}, nitrate assimilation; S_R, sulfur respiration; S_{DS}, dissimilation in sulfur metabolism; S_{AS}, assimilation in sulfur metabolism.

Constellations of the homologous proteins on the metabolic pathway map in the four energy metabolisms (photosynthesis, aerobic respiration, denitrification, and sulfur respiration) and other metabolisms (nitrate assimilation, dissimilation and assimilation in sulfur metabolism) are shown in Fig. 1. Black and blue arrows designate electron flow and compound change (enzymatic reaction), respectively. Squares are proteins. When one protein is composed by subunits, small squares showing subunits are shown in the lower part of the square. Only homologous proteins belonging to more than one metabolism system are shown in Fig. 1. Same-colored proteins are homologous. Because homology search was done by masking hydrophobic regions, the COX1 Pfam domain (group 25 in Fig. 1) was not detected in the present study.

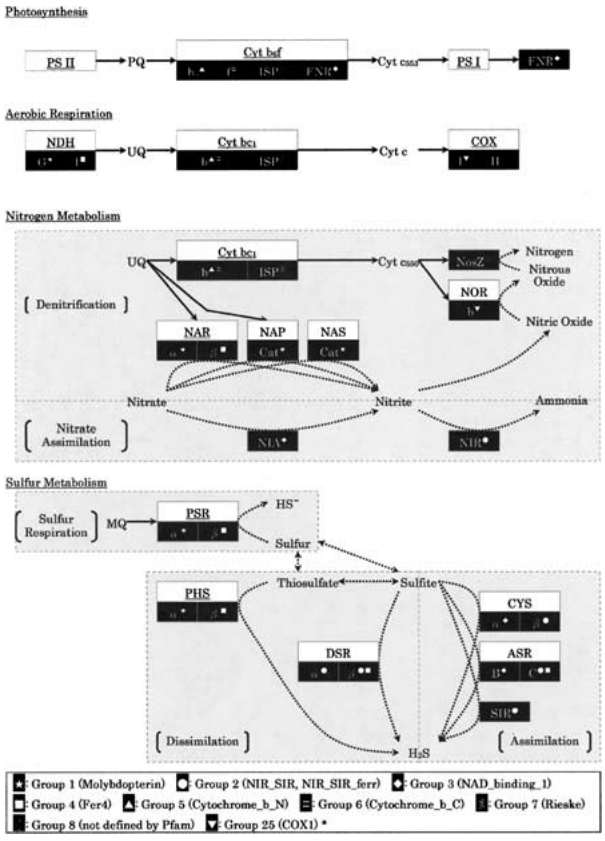

Fig. 1. Homologous proteins in the four metabolisms (photosynthesis, aerobic respiration, nitrogen metabolism, and sulfur metabolism). Contiguous and dotted arrows designate electron flow and compound change (enzymatic reaction), respectively. Squares are proteins. When one protein is composed of subunits, small squares showing subunits are shown in the lower part of the square. Some proteins involved in the four metabolisms are not shown in this figure. Proteins with same symbols are homologous. The explanatory notes of symbols are given at the bottom. Each symbol shows each homologous group. Group 25 proteins (COX1) were not detected by homology search analysis in this study, but they are thought to be homologous (Mogi et al. 1998; Saraste and Castresana 1994; Zumft et al. 1992). Subunits A to D and H to N of NDH are put together. Same alphabetical subunits of NDH in photosynthesis and NDH in oxidative phosphorylation are homologous. Conserved Pfam domains in each homologous group are shown in parentheses in the explanatory notes. PS II, photosystem II; NDH, NADH dehydrogenase; PQ, plastoquinone; Cyt b_6f, cytochrome b_6f complex; ISP, iron sulfur protein (subunit); FNR, ferredoxin NADP reductase; Cyt c_{553}, cytochrome c_{553} ; PS I, photosystem I; UQ, ubiquinone; Cyt bc_1, cytochrome bc_1 complex; Cyt c, cytochrome *c*; COX, cytochrome *c* oxidase; NAR, respiratory nitrate reductase; NAP, periplasmic nitrate reductase; NAS, assimilatory nitrate reductase; Cat, catalytic subunit; Cyt c_{550}, cytochrome c_{550} ; NIA, eukaryotic assimilatory nitrate reductase; NIR, nitrite reductase; NOR, nitric oxide reductase; NosZ, nitrous oxide reductase; MQ, menaquinone; PSR, polysulfide reductase; PHS, thiosulfate reductase; DSR, sulfite reductase, dissimilatory type; SIR, sulfite reductase (ferredoxin); CYS, sulfite reductase (NADPH); ASR, anaerobic sulfite reductase.

Group 1

Domain compositions of 44 proteins belonging to homologous group 1 are shown in Fig. 2. All group 1 proteins have the Molybdopterin domain, which has a binding site to molybdenum (Boyington et al. 1997; Czjzek et al. 1998; Bias et al. 1999; Jormakka et al. 2002; Li et al. 2000; McAlpine et al. 1998; Schindelin et al. 1996; Schneider et al. 1996; Stewart et al. 2000).

The phylogenetic tree of this group is presented in Fig. 3, based on the multiple alignment of the Molybdopterin domain. Although the nine sequences (FDXG_HAEIN, FDOG_ECOLI, FDNG_ECOLI, NQO3_THETH, NUG2_RHIME, NUOG_MYCTU, NUOG_ECOLI, NUOG_SALTY, and NUOG_STRCO) listed in Fig. 2 have the Molybdopterin domain, those domain sequences were fragmentary. We therefore excluded the nine sequences from the phylogenetic tree (Fig. 3). Topologies of the two NJ trees constructed using Poisson correction (not shown) and gamma distance (Fig. 3) were slightly different. We thus constructed a tree (not shown) using a combination of the Molybdopterin domain and Molydop_binding domain regions. Although nine proteins lack the Molydop_binding domain, they formed a clear monophyletic clusters with a high bootstrap probability (see Fig. 3), and exclusion of those sequences does not influence the overall tree topology. This new tree is more reliable than the tree constructed using only the Molybdopterin domain, with higher bootstrap probabilities (shown in parentheses in Fig. 3). The topology of this new tree was consistent with that of the tree computed for the gamma distance, thus we chose that tree as shown in Fig. 3.

All group 1 proteins are involved in enzymatic reactions accompanied by electron transfer. Group 1 proteins shown in Fig. 3 can be divided into five types based on domain structures corresponding to electron transfer patterns. Arrows in electron transfer patterns in Fig. 3 designate directions of electron flow. Electron flow in one pair of brackets occurs in each group 1 protein.

Type 1 proteins (NADH dehydrogenase subunit) do not bind molybdenum. Molybdenum plays an important role for reaction with substrates involving electron transfer. The main function of type 1 proteins is electron transfer from one subunit to another one through three ferredoxins; two 4Fe4Ss and one 2Fe2S (Ohnishi 1998). Only the type 1 protein has Fer2 domain that has 2Fe2S ferredoxin binding function.

Type 2 proteins (formate dehydrogenase subunit) bind molybdenum. Electron flow of this protein is as follows; electron donor (substrate [formate]) $\Longrightarrow$ molybdenum $\Longrightarrow$ 4Fe4S $\Longrightarrow$ electron acceptor (another subunit) (Boyington et al. 1997; Jormakka et al. 2002). The Molybdop_binding domain, found for most of the type 2 proteins, is important for binding molybdenum (Boyington et al. 1997; Czjzek et al. 1998; Dias et al. 1999; Jormakka et al. 2002; Schindelin et al. 1996; Schneider et al. 1996). FDHA_METJA lost this domain. The function of

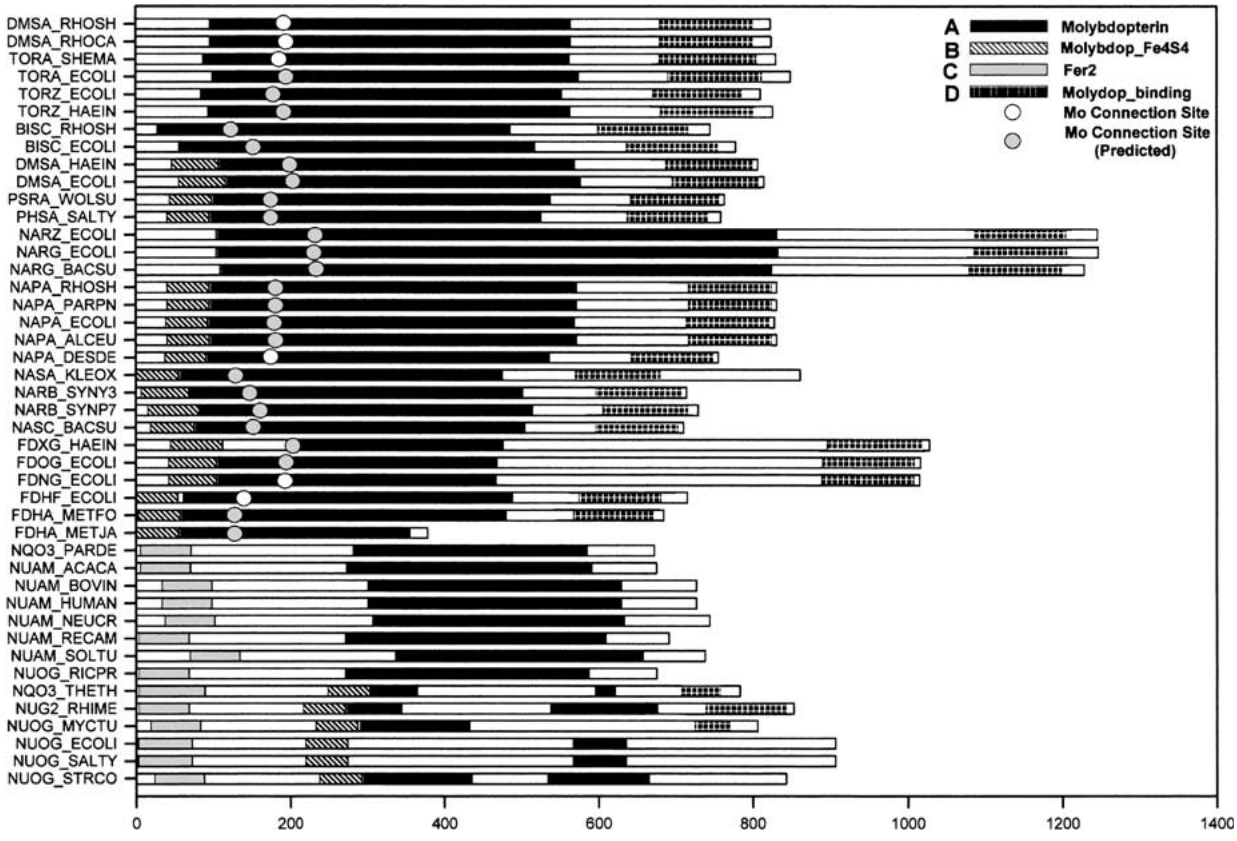

Fig. 2. Domain structures of group 1 proteins. SWISS-PROT IDs were used for protein names. Domains are defined following the Pfam database. Binding sites of molybdenum (Mo), which is important for catalysis, are shown by filled circles. We predicted molybdenum binding sites based on the multiple alignment of group 1 proteins. We regarded amino acids in the same column as known molybdenum bindings site as the molybdenum binding sites (shown by shaded circles).

FDHA_METJA was predicted only by sequence similarity (Table 2). FDHA_METJA may not bind molybdenum. Type 3 proteins (subunit of assimilatory and periplasmic nitrate reductase) have the same domains as type 2 proteins except for FDHA_METJA, but the electron flow is reversed (Breton et al. 1994; Butler et al. 1999; Dias et al. 1999). 4Fe4S ferredoxin is bound to the Molybdop_Fe4S4 domain, the 4Fe4S ferredoxin binding domain, in proteins of types 2 and 3.

Type 4 proteins (subunit of the trimer-type DMSO reductase and respiratory nitrate reductase) do not bind 4Fe4S ferredoxin. The Molybdop_4Fe4S domain of the trimer-type DMSO reductase subunit (DMSA_ECOLI and DMSA_HAEIN) is less conserved, and that of the respiratory nitrate reductase subunit (NARG_BACSU, NARG_ECOLI and NARZ_ECOLI) was lost. Four cysteines in this domain are essential for binding 4Fe4S ferredoxin. The arrangement of the four cysteines in type 4 is different from that of types 2 and 3 (Trieber et al. 1996). There are two amino acids between the first N-terminus two cysteines in type 2 and 3 proteins, but there are three amino acids in the trimer-type DMSO reductase subunit. In the respiratory nitrate reductase subunit, the first cysteine is replaced by histidine. Changed cysteine arrangement in type 4 lost the ability to bind ferredoxin (Magalon et al. 1998; Trieber et al. 1996). The cysteine arrangement is not found in type 5 proteins (biotin sulfoxide reductase, trimethylamine-*N*-oxide reductase, and monomer-type DMSO reductase), thus type 5 does not bind ferredoxin.

Although electron transfer patterns of types 4 and 5 look the same (see Fig. 3), electron transfer patterns among electron donor, molybdenum, and electron acceptor seem to be different between these two types. A type 4 protein gets an electron from the electron donor (substrate) at active site and probably gives it to the electron acceptor (another protein) by way of another side (Rothery et al. 1999). In contrast, a type 5 protein gives an electron to the electron acceptor at the same active site (Schindelin et al. 1996).

Which type of electron transfer pattern was ancestral among the five types in group 1 proteins? Type 2 proteins exist in eubacteria and archaea. This suggests that type 2 proteins emerged before the divergence of eubacteria and archaea. Types 3–5 proteins are found only in eubacteria (see Fig. 3), suggesting their relatively recent origin, after the divergence of eubacteria and archaea. Type 1 proteins exist in both

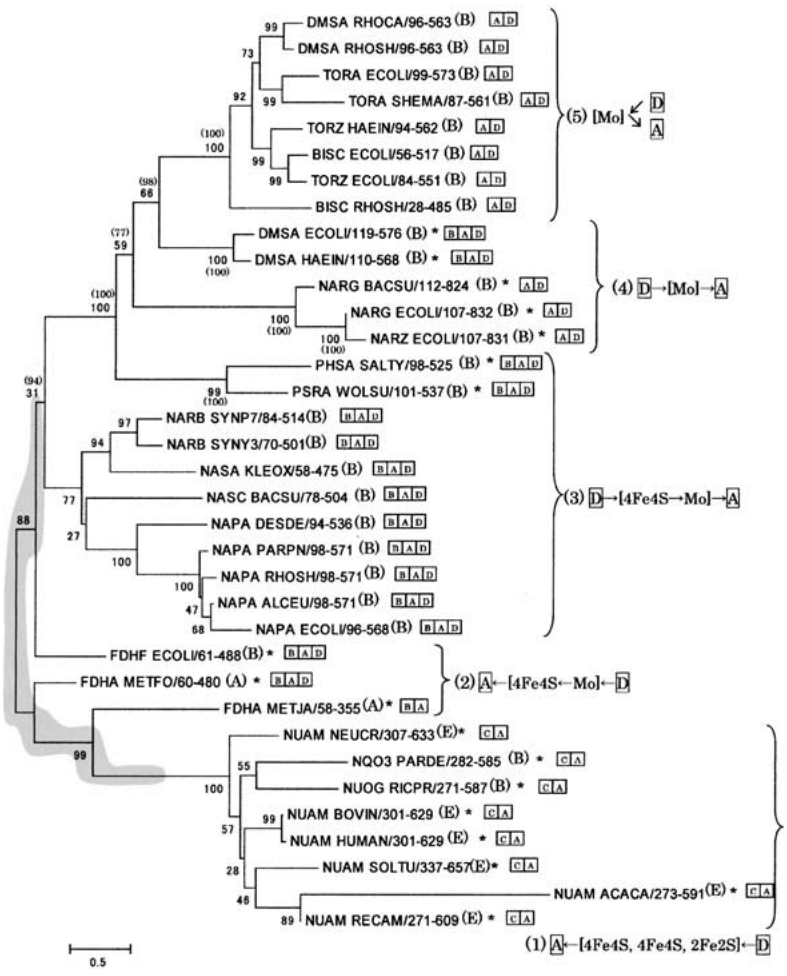

Fig. 3. A phylogenetic tree of group 1 proteins based on Molybdopterin-domain sequences. The shaded region shows possible locations of roots for this tree. Each square designates a Pfam domain. Explanatory notes on square symbols are given in Fig. 1. Electron transfer patterns are also shown. Arrows of the patterns indicate the direction of electron flow. D and A in squares: mean electron donor and acceptor, respectively. Unknown patterns of electron transfer were predicted from known patterns of evolutionarily closely related proteins. Proteins with an asterisk can generate a proton gradient.

eubacteria and eukaryotes. Type 1 protein genes in some eukaryote species (NUAM_ACACA and NUAM_RECAM) are encoded in their mitochondrial genomes. It is hypothesized that mitochondria was established as a result of endosymbiosis of an ancestor of the alpha proteobacterium *P. denitrificans*, and it is also hypothesized that the type 1 protein gene encoded in the mitochondrial genome was transferred to the eukaryotic genome after the endosymbiosis (Finel 1998). Andersson et al. (1998) showed that the alpha proteobacterium *R. prowazekii* is the closest eubacterial relative of mitochondria. Therefore, the type 1 protein gene of eukaryotes could be derived from that of the ancestor of alpha proteobacteria. Consequently, type 2 proteins seem to be the ancestor type protein in group 1. The gray-colored region in Fig. 3 shows the possible location of the root for this tree. We could not determine the exact position of the root.

Group 2

The domain structures of 21 proteins belonging to group 2 are shown in Fig. 4. All the group 2 proteins have the NIR_SIR Pfam domain and the NIR_SIR_ferr Pfam domain in adjacent locations. Both domains have siroheme binding sites. It should be noted that, confusingly, the NIR_SIR domain, not the NIR_SIR_ferr domain, have ferredoxin binding sites according to Pfam annotation (Crane et al. 1995). Hatched domains in Fig. 4 are not defined by Pfam, but they were detected as weakly similar to these two Pfam domains through PSI-BLAST search. Some of them were previously noticed by Crane and Getzoff (1996) and Larsen et al. (1999). We constructed a phylogenetic tree of group 2 proteins using the conserved regions consisting of the two domains (Fig. 5). The first nine proteins in Fig. 4 have two NIR_SIR_ferr and two NIR_SIR domains, although ferredoxin binding sites are lost in hatched NIR_SIR domains (Crane et al. 1995). These two NIR_SIR_ferr and two NIR_SIR domains suggest that an internal gene duplication occurred. We thus used each region as independent sequence in the case of these duplicated conserved regions. For example, we used two sequences in different conserved regions of CYSI_ECOLI; amino acid positions 109 to 233 and 387 to 489.

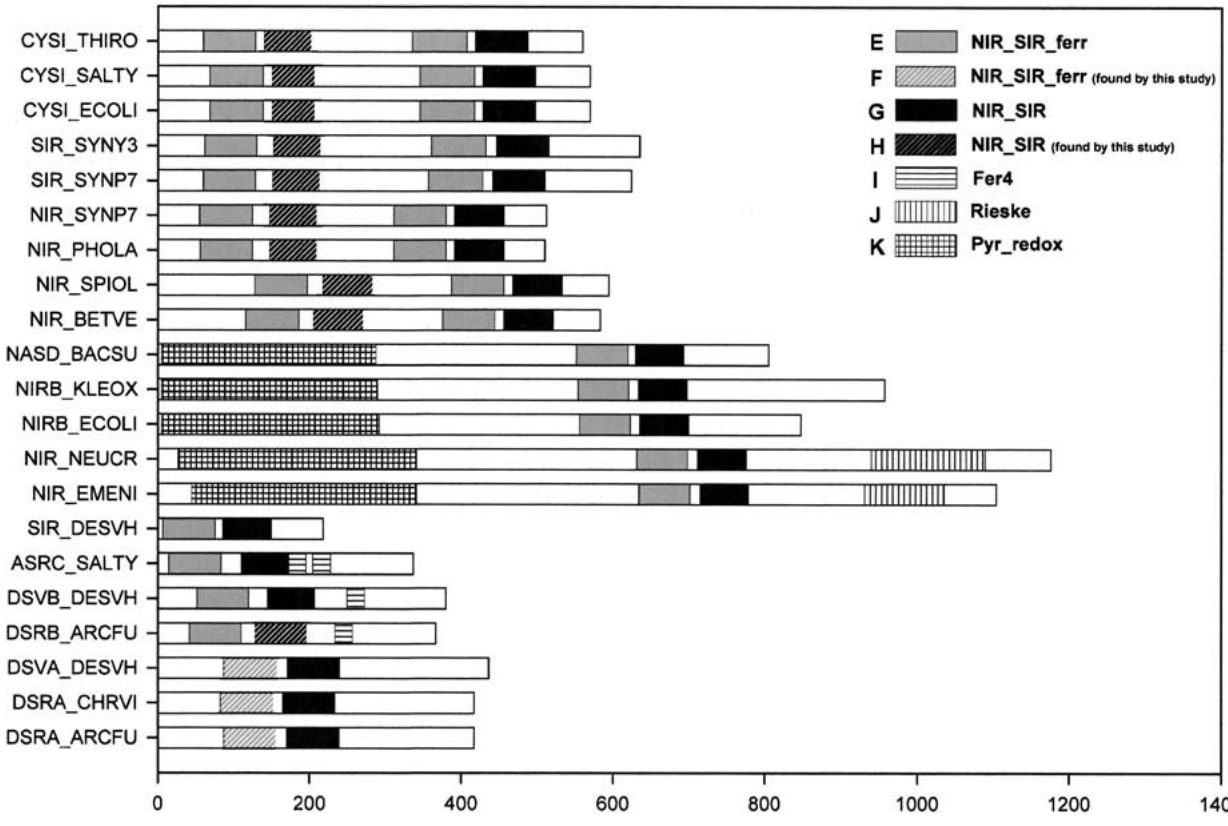

Fig. 4. Domain structures of group 2 proteins. SWISS-PROT IDs were used for protein names. Domains are defined following the Pfam database. Domains F and H were found in this study.

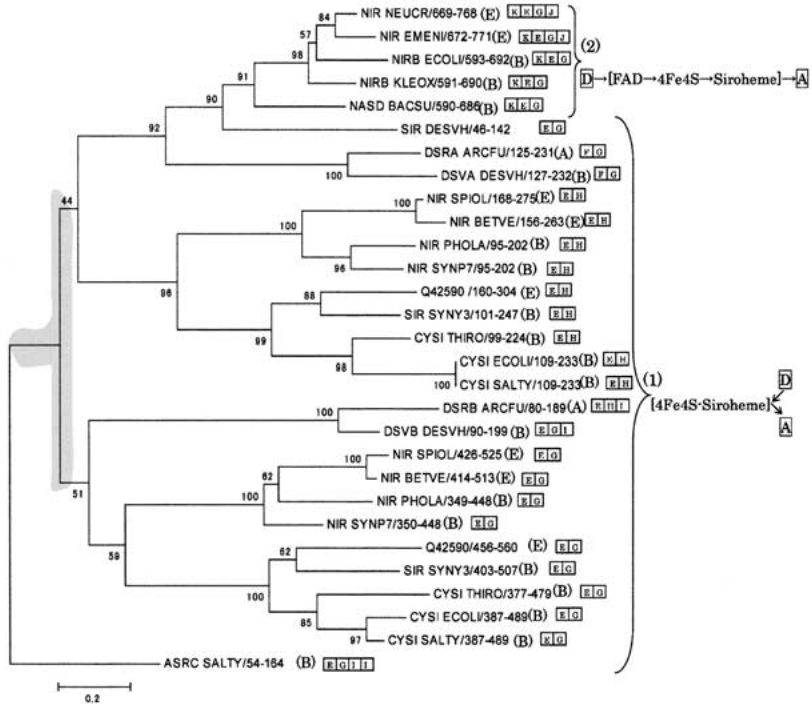

Fig. 5. A phylogenetic tree of group 2 proteins based on conserved regions consisting of the NIR_SIR_ferr domain and NIR_SIR domain. Conserved regions which are used for constructing this phylogenetic tree were obtained from previous works (Crane et al. 1997; Crane and Getzoff 1996; Larsen et al. 1999). Amino acid sequences of sulfite reductase (Q42590) from previous works are also added. The shaded region shows possible locations of roots for this tree. Each square in this figure shows a domain. Explanatory notes for each domain are given in Fig. 3. Arrows of electron transfer patterns indicate the direction of electron flow. D and A in squares: mean electron donor and acceptor, respectively. Unknown patterns of electron transfer were predicted from known patterns of evolutionarily closely related proteins.

Enzymatic reactions of group 2 proteins are accompanied by electron transfer. Arrows show electron flows. Group 2 proteins can be categorized into two types by the differences in electron transfer patterns (Fig. 5). Type 1 proteins bind siroheme and 4Fe4S ferredoxin, and electrons from the donor are transferred to the acceptor at these molecules (Crane et al. 1997; Tan and Cowan 1991). Type 2 proteins

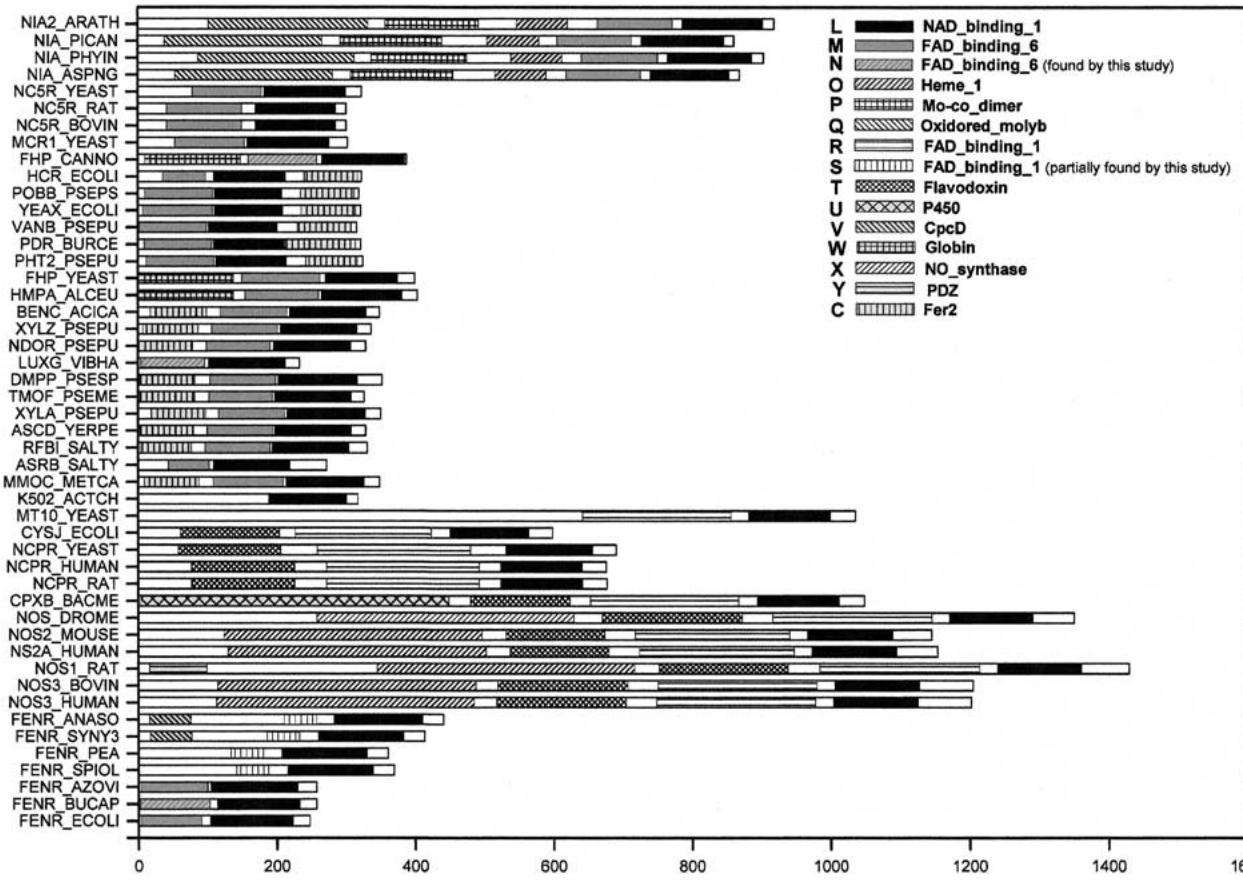

Fig. 6. Domain structures of group 3 proteins. SWISS-PROT IDs were used for protein names. Domains are defined following the Pfam database.

bind siroheme and 4Fe4S ferredoxin. However, electron flow is probably different from type 1 as follows; electron donor $\Longrightarrow$ FAD $\Longrightarrow$ 4Fe4S $\Longrightarrow$ siroheme $\Longrightarrow$ electron acceptor or substrate (Vega 1976). The tree topology of the phylogenetic tree shown in Fig. 5 suggests that the electron transfer pattern for type 1 is ancestral and that for type 2 is derived.

Larsen et al. (1999) noticed that the dissimilatory sulfite reductase alpha subunit (DSAR_ARCFU) and beta subunit (DSRB_ARCFU) of *Archaeoglobus fulgidus* may have been duplicated from an ancestral *dsr* gene before the divergence of eubacteria and archaea. The tree topology in Fig. 5 indicates that this gene duplication was followed by the emergence of proteins having a pair of NIR_SIR and NIR_-SIR_ferr domains. This suggests that duplicated genes were fused in the first nine proteins in Fig. 4. The root of the tree shown in Fig. 5 must be placed before this gene duplication. The gray-colored region in Fig. 5 thus shows the possible location of root for this tree.

Group 3

Homologous group 3 contains 122 proteins, and they were found using seven query sequences. The conserved region shared by all the proteins in this group was the NAD_binding_1 Pfam domain. The NAD_binding_1 domain has NAD(P)H connection sites. NAD(P)H works as electron donor or acceptor of the group 3 proteins. Because some proteins were closely related and formed clear monophyletic clusters in the phylogenetic tree when we used all proteins, we chose one amino acid sequence from each cluster. When more than one enzyme was found in one cluster, we chose representative proteins from each enzyme. A total of 48 proteins was thus used in the following analyses.

Domain structures of group 3 proteins are shown in Fig. 6, and they are all flavin enzymes which bind flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN). FAD is bound to the FAD_binding_6 domain or the FAD_binding_1 domain, and FMN is bound to the Flavodoxin or the FAD_binding_6 domain.

Enzymatic reactions of group 3 proteins are accompanied by electron transfer. Arrows in Fig. 7 are electron flows. Group 3 proteins can be categorized into 10 types based on electron transfer patterns (Fig. 7). Type 1 proteins (ferredoxin–NADP reductase) have a relatively simple domain structure, and both eukaryotes and eubacteria have

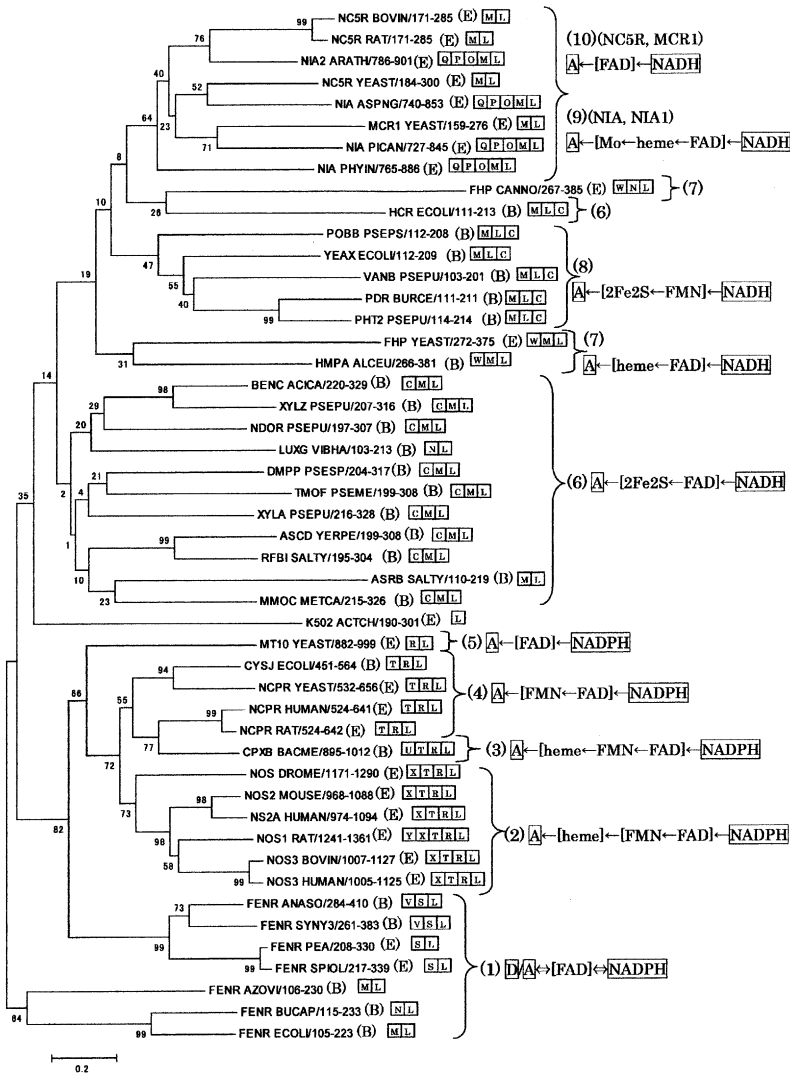

Fig. 7. A phylogenetic tree of group 3 proteins constructed based on NAD_binding_1 domain sequences. Each square shows a domain. Explanatory notes for each domain are given in Fig. 5. Arrows of electron transfer patterns indicate the direction of electron flow. D and A in squares mean electron donor and acceptor, respectively. Unknown patterns of electron transfer were predicted from known patterns of evolutionary closely related proteins.

this type. Furthermore, type 1 proteins are not monophyletic according to the tree topology (Fig. 7). Therefore, the common ancestor of group 3 proteins could have the type 1 electron transfer pattern.

Type 2 proteins (nitric oxide reductase) are homodimers. FAD, FMN, and heme are bound to each subunit. One subunit gets electron from NADPH and gives it to heme of the other subunit through FAD and FMN of the first subunit. Heme of the second subunit gives electron to electron acceptor. Electron from NADPH passes through two subunits and goes to electron acceptor. Electron flow in each pair of brackets shows electron flow occurring in each subunit.

Although type 3 proteins bind FAD, FMN, and heme as in the case of type 2 proteins, their electron transfer pattern is different from that of type 2 proteins. Type 3 proteins, are not homodimers. They receive an electron from NADPH and give it to the electron acceptor through FAD, FMN, and heme. The NO_synthase domain of type 2 proteins and the P450 domain of type 3 proteins have heme binding sites. This suggests that the insertion of the NO_synthase domain or the P450 domain enabled group 3 proteins to bind heme. Types 2, 3, and 4 proteins bind FAD and FMN. The FAD_binding_1 domain has FAD binding sites, and the Flavodoxin domain has FMN binding sites. Insertion of the Flavodoxin domain may have enabled group 3 proteins to bind FMN.

Figure 8A shows the phylogenetic tree of the proteins belonging to types 2–5, constructed based on the FAD_binding_6 and NAD_binding_1 domains. The topology of this tree is more reliable than that of the tree constructed based on only NAD_binding_1 domain sequences. Type 5 proteins do not have the Flavodoxin domain. Therefore, the tree in Fig. 8A suggests that insertion of the Flavodoxin domain occurred on the lineage going to types 2, 3, and 4 proteins. Figure 8A also suggests that NO_synthase and P450 domains that have binding sites of heme

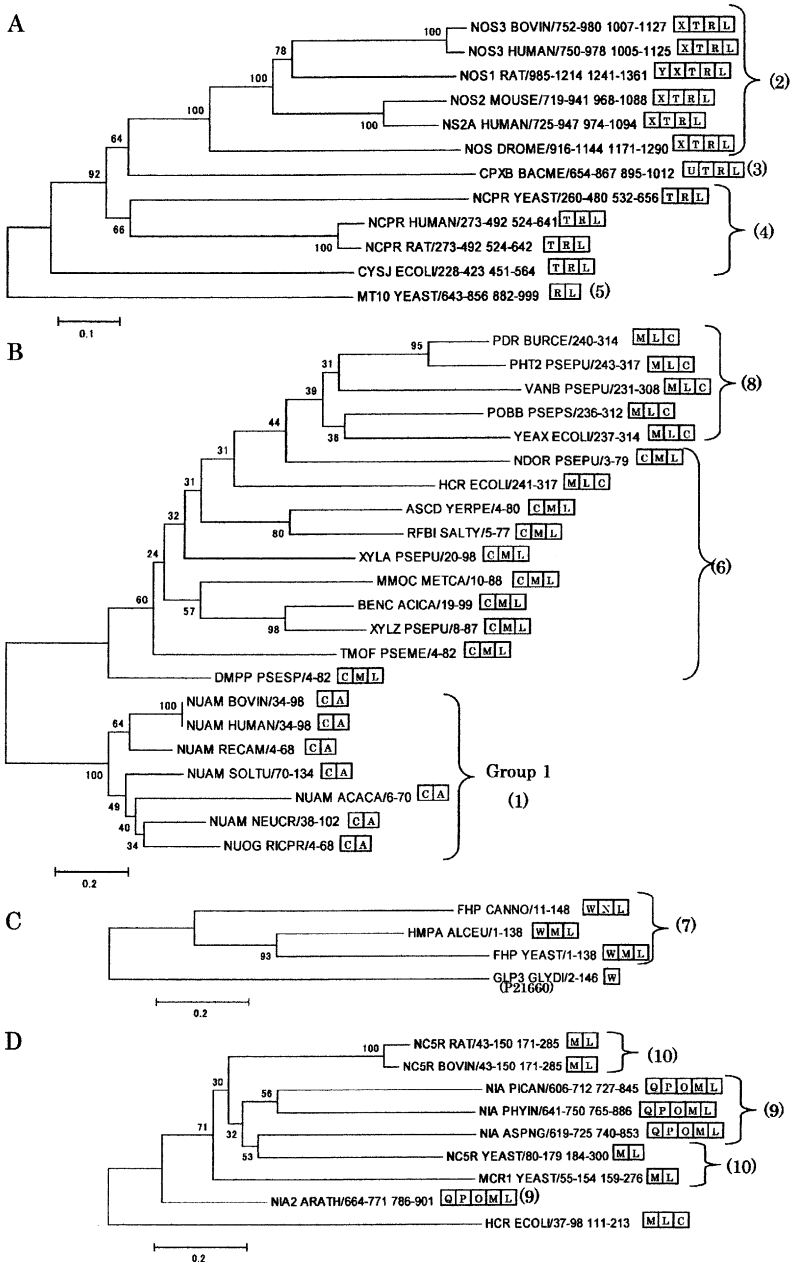

Fig. 8. **A** A phylogenetic tree of the type 2, 3, 4, and 5 proteins based on the FAD_binding_1 and NAD_binding_1 domains. **B** A phylogenetic tree of the type 1 proteins in group 1 and the type 6 and 8 proteins in group 3 based on the Fer2 domain. **C** A phylogenetic tree of the type 7 proteins constructed based on the Globin domain. **D** A phylogenetic tree of the type 9 and 10 proteins based on the FAD_binding_6 and NAD_binding_1 domains.

were inserted on the lineage to type 2 and type 3 proteins, respectively.

Types 6 and 8 proteins both bind 2Fe2S ferredoxin. Ferredoxin is bound to the Fer2 domain. This suggests that insertion of the Fer2 domain enabled group 3 proteins to bind ferredoxin. Figure 8B shows the phylogenetic tree constructed based on the Fer2 domain amino acid sequences. Not only types 6 and 8 of group 3 but also type 1 proteins of group 1 have this domain. Proteins of this tree can be largely categorized into two lineages; type 1 proteins in group 1 and types 6 and 8 proteins in group 3. The tree topology of group 3 proteins in Fig. 8B suggests that the Fer2 domain was first inserted to form type 6 proteins and was further inserted to form type 8 proteins. The inserted location of the Fer2 domain is different between type 6 and type 8 proteins except for HCR_ECOLI (Figs. 5 and 8B). Although the domain structure of HCR_ECOLI is similar to that of type 8 proteins, the electron transfer pattern of HCR_ECOLI is similar to that of type 6 proteins. HCR_ECOLI is phylogenetically closer to type 6 proteins (see Fig. 8B). One possibility is that domain inversion occurred in the HCR_ECOLI lineage.

Hemes in type 7 proteins are bound to the Globin domain (see Fig. 7). The insertion of the Globin domain may have enabled group 3 proteins to bind heme. Type 7 proteins are not monophyletic;

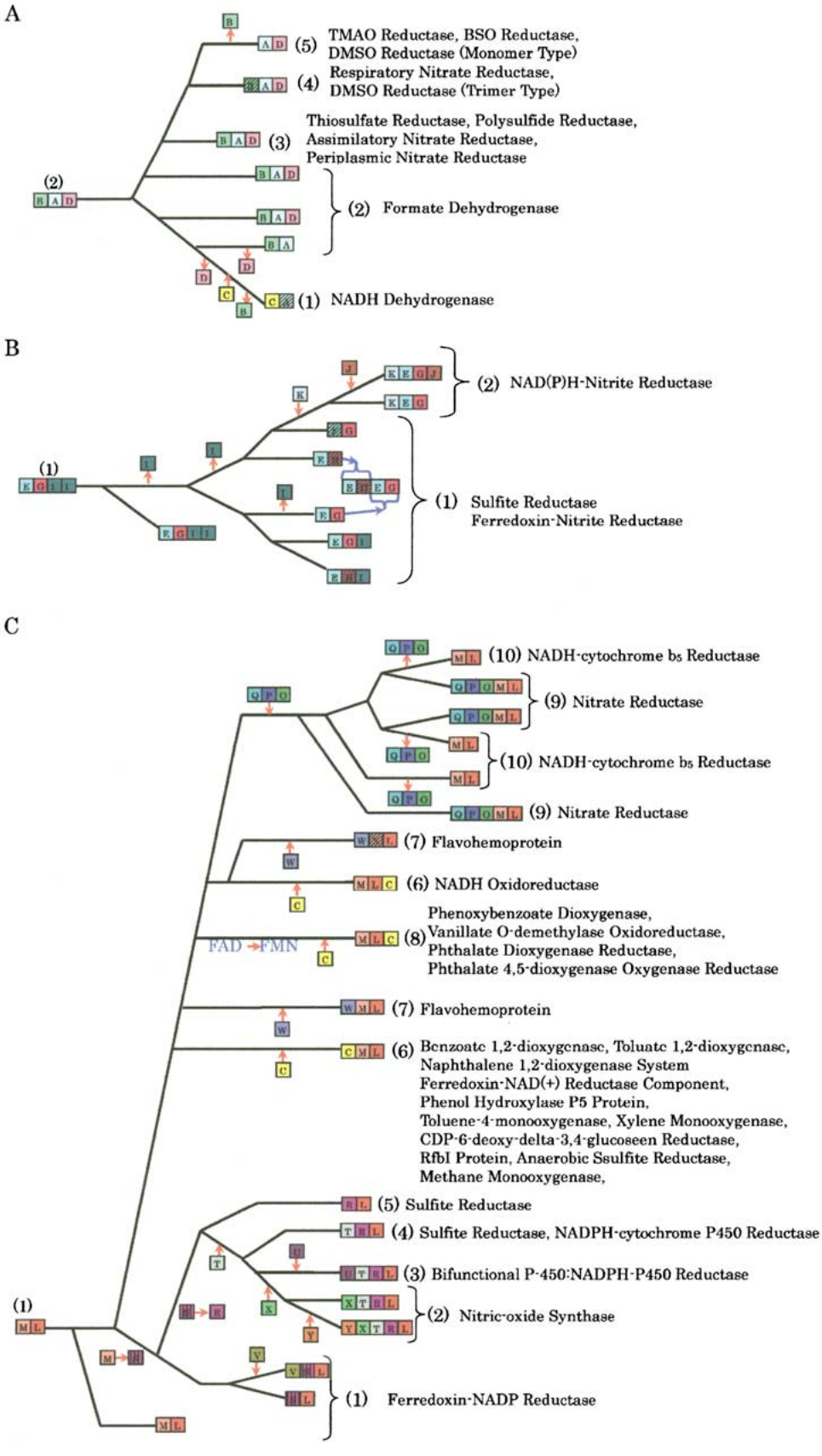

Fig. 9. Estimated domain insertion/deletion events during evolution of the three protein groups. Each square shows a domain defined by Pfam. **A** Group 1 proteins. Square symbols are the same as in Fig. 1. **B** Group 2 proteins. The correspondence between square symbols and Pfam domain is shown in Fig. 3. **C** Group 3 proteins. Square symbol of each domain is the same as in Fig. 5.

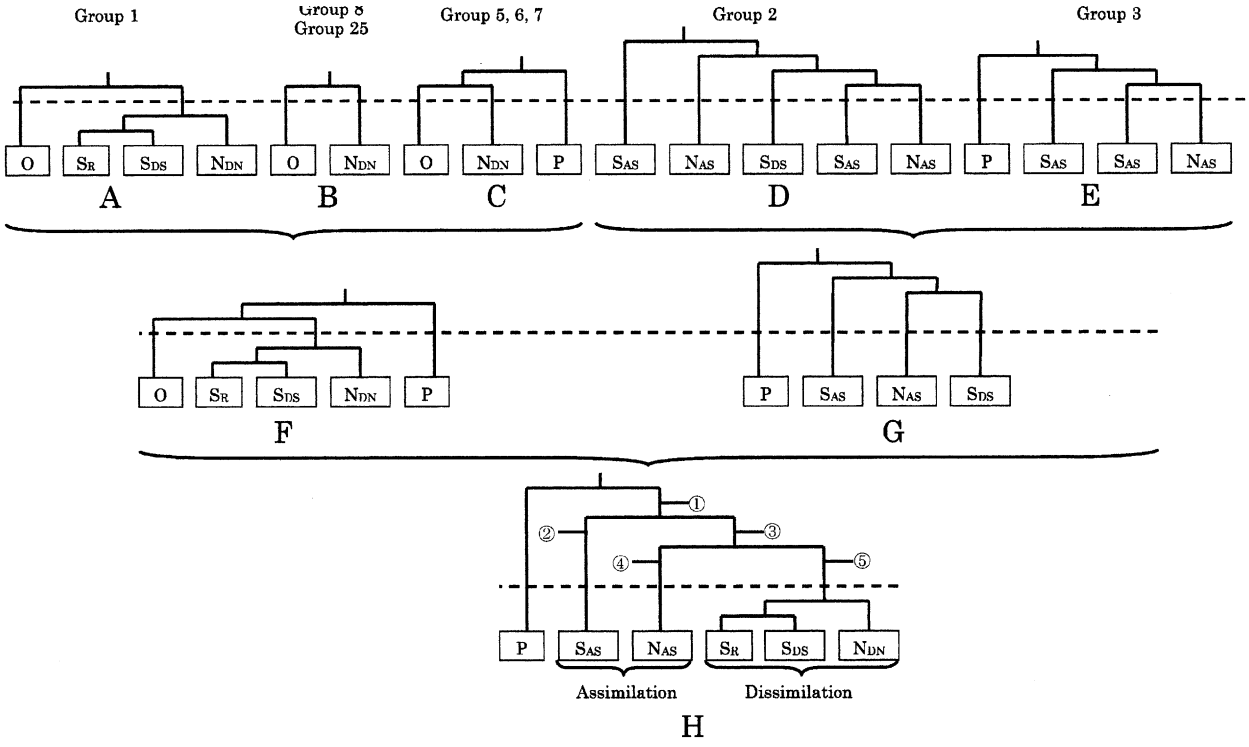

Fig. 10. Phylogenetic trees of proteins shown in Fig. 9 (upper trees). Branch length has no meaning. The dashed lines indicate divergence between eubacteria and others (archaea and eukaryotes). Divergences between eubacteria and others were estimated in this study (Figs. 2, 4, and 6) and previous studies (Berry et al. 2000; Saraste and Castresana 1994). The lower trees are superimposed trees of the upper trees. Each OTU name is the metabolism in which each protein works: P, photosynthesis; O, aerobic respiration; N_{DN}, denitrification; N_{AS}, nitrate assimilation; S_R = sulfur respiration; S_{DS}, dissimilation in sulfur metabolism; S_{AS}, assimilation in sulfur metabolism.

HMPA_ALCEU (bacteria) and FHP_YEAST (eukaryote) formed a weak monophyletic cluster (the bootstrap probability is only 30%), while FHP_CANNO is located in a different lineage (see Fig. 7). Figure 8C shows the phylogenetic tree of type 7 proteins constructed based on the Globin-domain amino acid sequences. This tree also suggests that HMPA_ALCEU and FHP_YEAST form a monophyletic cluster. The bootstrap probability of this cluster is now 93%. Therefore, it is possible that the lineage of HMPA_ALCEU and FHP_YEAST is separated from that of FHP_YEAST. This implies that the Globin domain may have been inserted independently on the lineage to HMPA_ALCEU and FHP_YEAST and the lineage to FHP_YEAST.

Type 9 proteins are nitrate reductase, and they bind FAD according to the tertiary structure analysis (Lu et al. 1994; 1995). It was hypothesized by sequence similarity that heme and molybdenum are bound to the Heme_1 domain and Oxidored_molyb domain of type 9 protein, respectively (Avila et al. 1995; Crawford et al. 1988; Pieterse et al. 1995; Unkles et al. 1992). This suggests that insertion of the Oxidored_molyb domain enabled group 3 proteins to bind molybdenum, and insertion of the Heme_1 domain enabled group 3 proteins to bind heme. Figure 8D shows the phylogenetic tree of type 9 and type 10 proteins constructed based on the NAD_binding_1 and FAD_binding_6 domains. The topology of this tree is more reliable than that of the tree in Fig. 7, which was constructed based on only NAD_binding_1 domain sequences. This tree suggests that type 9 and 10 proteins are not monophyletic. However, bootstrap probabilities of various branches are not high, so it is not clear.

Discussion

Domain Change and Electron Transfer Pattern Change Through Evolution

The phylogenetic analysis of the three groups suggests that each homologous group experienced frequent insertions and deletions of domains. We thus estimated domain insertion/deletion events during evolution and connected those events with electron transfer pattern changes.

Group 1

We hypothesized the evolutionary process of group 1 proteins, as depicted in Fig. 9A, based on the phylogenetic tree shown in Fig. 3. We assumed that type 2 was the ancestral type of electron transfer pattern, and

the domain structure of the ancestral protein was assumed to be the same as that of type 2 proteins. Type 2 proteins have Molybdop_Fe4S4 and Molybdopterin domains. Two of the type 2 proteins have the Molydop_binding domain, which is important for binding molybdenum, but FDHA_METJA does not have this domain. As we described above, we suspected that FDHA_METJA does not bind molybdenum and lost the catalytic activity. Because some NADH dehydrogenase subunits (type 1 proteins) have the Molydop_biding domain (Fig. 2), it seems that the ancestral protein also had this domain. Therefore, we assumed that the Molydop_binding domain was lost independently in the lineage to FDHA_METJA and on the type 1 proteins (Fig. 9A). If this hypothesis is true, the ancestral protein probably had the Molydop_Fe4S4, Molybdopterin, and Molydop_binding domains as is the case for the type 2 proteins except for FDHA_METJA. The ancestral protein may have bound molybdenum and 4Fe4S ferredoxin.

A scenario of evolutionary changes of the group 1 proteins from this ancestral protein is as follows. On the lineage to type 1 proteins (NADH dehydrogenase), the Fer2 domain was acquired, and ferredoxin was bound to this domain. In contrast, the Molybdopterin domain became less conserved, and most type 1 proteins lost the Molydop_binding domain. Both the Fer2 and the ferredoxin domains are important for molybdenum binding. Therefore type 1 proteins lost the function of binding molybdenum, which is important for catalysis. Their main function changed from catalysis to electron transfer. On the lineage to type 3, the direction of electron flow was reversed. This reversal did not seem to be caused by domain composition changes. Instead, amino acid substitutions for changing reactant specificities probably occurred on the lineage to type 3, and the changing of reactants caused the reversal of electron flow direction. The molybdop_4Fe4S domain that has binding sites of 4Fe4S ferredoxin became less conserved on the lineage to type 4 proteins and was completely lost on the lineage to type 5. Types 4 and 5 proteins lost the function of binding 4Fe4S ferredoxin. In type 5 proteins, electron is exchanged at molybdenum from electron donor to electron acceptor. There are no other electron carriers in this protein.

The common feature of the group 1 proteins is that all of them have the Molybdopterin domain (Fig. 2), although this domain is not conserved well in type 1 proteins. Other features are not necessarily common. For example, ferredoxin is attached to types 1–3 proteins but not to others (Fig. 3). Molybdenum is attached to types 2–5 proteins but not to type 1 (Fig. 3). Some group 1 proteins (indicated by asterisks in Fig. 3) are subunits of proteins that can generate the proton gradient. The proton gradient is the energy source for generating ATP by ATP synthase. Based on our phylogenetic analysis, this important function, proton gradient generation, is not necessarily conserved. Although the general electron transfer patterns are more or less similar in type 3 proteins, some proteins (polysulfide reductase subunit and thiosulfate reductase subunit) work for proton gradient generation, but the others (assimilatory nitrate reductase subunit) do not (see Fig. 3).

Group 2

We conjectured the evolution of group 2 proteins as shown in Fig. 9B. The common ancestral protein was assumed to have both the NIR_SIR_ferr and the NIR_SIR domains, because all group 2 proteins have these two domains including weakly conserved ones. The two Fer4 domains were assumed to exist in the common ancestral protein, as in the case of the ASRC_SALTY protein. Under this scenario, three losses of the Fer4 domain were assumed (Fig. 9B). However, proteins with only one or no Fer4 domain could also be the common ancestor. All three possibilities are equally parsimonious for the three insertion/deletion events of the Fer4 domain.

Later, a gene duplication produced two copies of the NIR_SIR_ferr and NIR_SIR domains. This means that one gene including the NIR_SIR_ferr and NIR_SIR domains was duplicated, and two genes like a pair of DSRA (alpha subunit of sulfite reductase) and DSRB (beta subunit), or DSVA (alpha subunit of sulfite reductase) and DSVB (beta subunit), were produced. These two genes were later fused. In the lineage of DSRA and DSVA, the Pyr_redox domain was added and the FAD binding site was generated. This creation of the FAD binding capacity seems to be the critical point for generating type 2 proteins.

Group 3

A parsimoniously hypothesized evolutionary history of group 3 proteins is shown in Fig. 9C. As we described in the Results, the electron transfer pattern of the ancestral proteins seems to have been type 1. If so, that ancestral protein had NAD_binding_1 and FAD_binding_6 domains like type 1 proteins of eubacteria.

On the lineage to types 2–4, the Flavodoxin domain having the FMN binding region was inserted. Because of this addition, FMN joined as a member of electron flow. Domains of P450 and NO_synthase that have binding sites for heme were inserted on the lineages to type 2 and type 3 proteins, respectively. Types 2 and 3 proteins thus acquired the ability to bind heme. On the lineages to types 6 and 8, 2Fe2S ferredoxin was added by

the insertion of the Fer2 domain, which has 2Fe2S ferredoxin binding sites. FAD was replaced by FMN in type 8, although the FAD_binding_6 domain, which has the binding region of FAD or FMN, did not change. Heme was also added on the lineage to type 7 proteins because of the insertion of the Globin domain, which has heme binding sites.

Lineages to types 9 and 10 proteins are more complicated. Based on the phylogenetic tree in Fig. 8D, we hypothesized that the insertion of the three domains (Oxidored_molyb, Mo-co_dimer, and Heme_1) had occurred only once. The three domains were then deleted three times on the lineage to MCR1_YEAST, the lineage to NC5R_YEAST, and the lineage to NC5R_BOVIN and NC5R_RAT independently. But the branching pattern of types 9 and 10 is not supported by high bootstrap probabilities. Therefore, we cannot predict how these proteins were evolved in detail. In Fig. 9C, we illustrate the lineages of type 9 and 10 proteins that we hypothesized based on Fig. 8D.

The common features of the group 3 proteins are that all proteins have the NAD_binding_1 domain. NAD(P)H is their electron donor or acceptor, and FAD or FMN is bound to them as an electron transfer component. Although these features are conserved, electron transfer patterns in group 3 proteins are so diverged by adding and deleting various domains.

Evolution of the Metabolism Systems

The four energy metabolisms and dissimilation in sulfur metabolism are energy generation metabolisms. Proteins which can generate proton potential (abbreviations of such proteins are underlined in Fig. 1) are conserved between different metabolism systems. Proton potential is the source for the synthesis of ATP by ATP synthase. This function is important for energy generation metabolisms. Proteins which have this important function, generation of proton potential, seem to be conserved among the four energy metabolisms and dissimilation in sulfur metabolism. This suggests that those systems are evolutionary related.

We therefore tried to infer the evolutionary relationship of the four energy metabolisms and dissimilation metabolism. Trees A–E in Fig. 10 show the five phylogenetic trees of homologous proteins analyzed in this study and previous analyses (Berry et al. 2000; Saraste et al. 1994). Divergence between eubacteria and others (archaea and eukaryotes) are shown in Fig. 10 based on rough estimations of this study (Figs. 3, 5, and 7) and previous studies (Berry et al. 2000; Saraste and Castresana 1994).

Those protein trees were superimposed to infer the evolutionary tree of metabolic systems. The idea of superimposition came from Oota and Saitou (1999), who inferred the evolutionary tree of muscle tissues by superimposition of muscle protein trees. Because the protein cluster of denitrification (N_{DN}), sulfur respiration (S_R), and dissimilation in sulfur metabolism (S_{DS}) was estimated to emerge after the divergence of eubacteria and archaea/eukaryotes according to tree A, these three systems were considered to be closely related. Therefore, N_{DN} of trees B and C in Fig. 10 were assumed to correspond to the N_{DN} –S_R –S_{DS} cluster of tree A. Trees A, B, and C in Fig. 10 were thus superimposed to produce tree F.

We then superimposed trees D and E in Fig. 10. Because our interest is the order of system generation, we ignored the newer lineage of nitrate assimilation (N_{AS}) and assimilation in sulfur metabolism (S_{AS}) in these two trees. The resultant superimposed tree is tree G in Fig. 10. Trees F and G were further superimposed to produce single phylogenetic tree H of the seven metabolic systems. However, because of the lack of information, the branching point of the aerobic respiration system was not able to be determined, and there are five branching possibilities (1–5 in tree H, Fig. 10). Tree H suggests that photosynthesis is basal (diverged first) among the energy metabolisms we analyzed. Aerobic respiration might be utilized after the increase in oxygen by photosynthesis, while ancestral sulfur and/or nitrogen metabolism system coexisted with that. Therefore, we considered that branching position 1 may be most plausible among the five possibilities for the emergence of the aerobic respiration system (see tree H in Fig. 10). In fact, a recent study suggested that the atmospheric oxygen was present at significant levels around 2.3 billion years ago (Bekker et al. 2004).

If we accept tree H, the assimilation process (nitrate assimilation and assimilation in sulfur metabolism) seems to be older than the dissimilation process (denitrification, sulfur respiration, and dissimilation in sulfur metabolism). Assimilation systems might be generated before the divergence of eubacteria and archaea/eukaryotes, and dissimilation systems might be generated after the divergence of eubacteria and archaea/eukaryotes.

Although tree H does not show the ancestral system at the branching nodes, we can infer that nitrate assimilation was ancestral to the sulfur assimilation metabolism. Nitrate assimilation is involved in the metabolism for producing any amino acids, while assimilation in sulfur metabolism is involved only for producing amino acids including sulfur (cysteine and methionine). It is conceivable that generation of nitrate assimilation enabled ancient organisms to pro-

duce amino acids except for cysteine and methionine at first, followed by acquisition of the ability for producing cysteine and methionine by assimilation in sulfur metabolism.

The above conjectures were based on the superimposed tree H in Fig. 10. Because this tree was produced through various assumptions, some or all of our conjectures may turn out to be incorrect. Nonetheless, we would like to point out the possibility that the phylogenetic tree of proteins can infer the phylogenetic relationship of metabolism systems.

Acknowledgments. We are grateful to Dr. Hiromi Suzuki of Meiji University for helping us in the initial part of this study. We appreciate Dr. Takeshi Kawabata of NAIST for advising us of the PSI-BLAST homology search method and helping us to conduct it. Color version of all figures is available at http://sayer.lab.nig.ac.jp/~tomiki/JME_metabolism/.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396:133–140

Avila J, Perez MD, Brito N, Gonzalez C, Siverio JM (1995) Cloning and disruption of the YNR1 gene encoding the nitrate reductase apoenzyme of the yeast *Hansenula polymorpha*. FEBS Lett 366:137–142

Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL (2002) The Pfam protein families database. Nucleic Acids Res 30:276–280

Bekker A, Holland HD, Wang PL, Rumble III D, Stein HJ, Hannah JL, Coetzee LL, Beukes NJ (2004) Dating the rise of atmospheric oxygen. Nature 427:117–120

Berry EA, Guergova-Kuras M, Huang LS, Crofts AR (2000) Structure and function of cytochrome be complexes. Annu Rev Biochem 69:1005–1075

Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M (2003) The SWISS-PROT protein knowledge base and its supplement TrEMBL in 2003. Nucleic Acids Res 31:365–370

Boyington JC, Gladyshev VN, Khangulov SV, Stadtman TC, Sun PD (1997) Crystal structure of formate dehydrogenase H: catalysis involving Mo, molybdopterin, selenocysteine, and an Fe_4S_4 cluster. Science 275:1305–1308

Breton J, Berks BC, Reilly A, Thomson AJ, Ferguson SJ, Richardson DJ (1994) Characterization of the paramagnetic iron-containing redox centres of *Thiosphaera pantotropha* periplasmic nitrate reductase. FEBS Lett 345:76–80

Butler CS, Charnock JM, Bennett B, Sears HJ, Reilly AJ, Ferguson SJ, Garner CD, Lowe DJ, Thomson AJ, Berks BC, Richardson DJ (1999) Models for molybdenum coordination during the catalytic cycle of periplasmic nitrate reductase from *Paracoccus denitrificans* derived from EPR and EXAFS spectroscopy. Biochemistry 38:9000–9012

Crane BR, Getzoff ED (1996) The relationship between structure and function for the sulfite reductases. Curr Opin Struct Biol 6:744–756

Crane BR, Siegel LM, Getzoff ED (1995) Sulfite reductase structure at 1.6 A: Evolution and catalysis for reduction of inorganic anions. Science 270:59–67

Crane BR, Siegel LM, Getzoff ED (1997) Probing the catalytic mechanism of sulfite reductase by X-ray crystallography: Structures of the *Escherichia coli* hemoprotein in complex with substrates, inhibitors, intermediates, and products. Biochemistry 36:12120–12137

Crawford NM, Smith M, Bellissimo D, Davis RW (1988) Sequence and nitrate regulation of the *Arabidopsis thaliana* mRNA encoding nitrate reductase, a metalloflavoprotein with three functional domains. Proc Natl Acad Sci USA 85:5006–5010

Czjzek M, Dos Santos JP, Pommier J, Giordano G, Mejean V, Haser R (1998) Crystal structure of oxidized trimethylamine N-oxide reductase from *Shewanella massilia* at 2.5 A resolution. J Mol Biol 284:435–447

Dias JM, Than ME, Humm A, Huber R, Bourenkov GP, Bartunik HD, Bursakov S, Calvete J, Caldeira J, Carneiro C, Moura JJ, Moura I, Romao MJ (1999) Crystal structure of the first dissimilatory nitrate reductase at 1.9 A solved by MAD methods. Struct Fold Des 7:65–79

Finel M (1998) Organization and evolution of structural elements within complex I. Biochim Biophys Acta 1364:112–121

Hauska G, Nitschke W, Herrmann RG (1988) Amino acid identities in the three redox center carrying polypeptides of cytochrome bc_1/b_6 f complexes. J Bioenerg Biomembr 20:211–228

Jormakka M, Tornroth S, Byrne B, Iwata S (2002) Molecular basis of proton motive force generation: Structure of formate dehydrogenase-N. Science 295:1863–1868

Kanehisa M, Goto S, Kawashima S, Nakaya A (2002) The KEGG databases at GenomeNet. Nucleic Acids Res 30:42–46

Krafft T, Bokranz M, Klimmek O, Schroder I, Fahrenholz F, Kojro E, Kroger A (1992) Cloning and nucleotide sequence of the *psrA* gene of *Wolinella succinogenes* polysulphide reductase. Eur J Biochem 206:503–510

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular evolutionary genetics analysis software. Bioinformatics 17:1244–1245

Larsen O, Lien T, Birkeland NK (1999) Dissimilatory sulfite reductase from *Archaeoglobus profundus* and *Desulfotomaculum thermocisternum*: Phylogenetic and structural implications from gene sequences. Extremophiles 3:63–70

Li HK, Temple C, Rajagopalan KV, Schindelin H (2000) The 1.3 A crystal structure of *Rhodobacter sphaeroides* dimethyl sulfoxide reveals two distinct molybdenum coordination environments. J Am Chem Soc 122:7673–7680

Lu G, Campbell WH, Schneider G, Lindqvist Y (1994) Crystal structure of the FAD-containing fragment of corn nitrate reductase at 2.5 A resolution: Relationship to other flavoprotein reductases. Structure 2:809–821

Lu G, Lindqvist Y, Schneider G, Dwivedi U, Campbell W (1995) Structural studies on corn nitrate reductase: refined structure of the cytochrome b reductase fragment at 2.5 A, its ADP complex and an active-site mutant and modeling of the cytochrome b domain. J Mol Biol 248:931–948

Magalon A, Asso M, Guigliarelli B, Rothery RA, Bertrand P, Giordano G, Blasco F (1998) Molybdenum cofactor properties and [Fe-S] cluster coordination in *Escherichia coli* nitrate reductase A: investigation by site-directed mutagenesis of the conserved his-50 residue in the NarG subunit. Biochemistry 37:7363–7370

McAlpine AS, McEwan AG, Bailey S (1998) The high resolution crystal structure of DMSO reductase in complex with DMSO. J Mol Biol 275:613–623

Mogi T, Tsubaki M, Hori H, Miyoshi H, Nakamura H, Anraku Y (1998) Two terminal quinol oxidase families in *Escherichia coli*: Variations on molecular machinery for dioxygen reduction. J Biochem Mol Biol Biophys 2:79–110

Moreno-Vivian C, Cabello P, Martinez-Luque M, Blasco R, Castillo F (1999) Prokaryotic nitrate reduction: Molecular properties and functional distinction among bacterial nitrate reductases. J Bacteriol 181:6573–6584

Ohnishi T (1998) Iron–sulfur clusters/semiquinones in complex I. Biochim Biophys Acta 1364:186–206

Ota S, Saitou N (1999) Phylogenetic relationship of muscle tissues deduced from superimposition of gene trees. Mol Biol Evol 16:856–867

Park J, Karplus K, Barrett C, Hughey R, Haussler D, Hubbard T, Chothia C (1998) Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods. J Mol Biol 284:1201–1210

Pieterse CM, van't Klooster J, van den Berg-Velthuis GC, Govers F (1995) *NiaA*, the structural nitrate reductase gene of *Phytophthora infestans*: Isolation, characterization and expression analysis in *Aspergillus nidulans*. Curr Genet 27:359–366

Rothery RA, Trieber CA, Weiner JH (1999) Interactions between the molybdenum cofactor and iron-sulfur clusters of *Escherichia coli* dimethylsulfoxide reductase. J Biol Chem 274:13002–13009

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Saraste M, Castresana J (1994) Cytochrome oxidase evolved by tinkering with denitrification enzymes. FEBS Lett 341:1–4

Schindelin H, Kisker C, Hilton J, Rajagopalan KV, Rees DC (1996) Crystal structure of DMSO reductase: Redox-linked changes in molybdopterin coordination. Science 272:1615–1621

Schneider F, Lowe J, Huber R, Schindelin H, Kisker C, Knablein J (1996) Crystal structure of dimethyl sulfoxide reductase from *Rhodobacter capsulatus* at 1.88 A resolution. J Mol Biol 263:53–69

Stewart LJ, Bailey S, Bennett B, Charnock JM, Garner CD, McAlpine AS (2000) Dimethylsulfoxide reductase: An enzyme capable of catalysis with either molybdenum or tungsten at the active site. J Mol Biol 299:593–600

Tan J, Cowan JA (1991) Enzymatic redox chemistry: A proposed reaction pathway for the six-electron reduction of SO_3^{2-} to S^{2-} by the assimilatory-type sulfite reductase from *Desulfovibrio vulgaris* (Hildenborough). Biochemistry 30:8910–8917

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Trieber CA, Rothery RA, Weiner JH (1996) Engineering a novel iron-sulfur cluster into the catalytic subunit of *Escherichia coli* dimethyl-sulfoxide reductase. J Biol Chem 271:4620–4626

Unkles SE, Campbell EI, Punt PJ, Hawker KL, Contreras R, Hawkins AR, Van den Hondel CA, Kinghorn JR (1992) The *Aspergillus niger niaD* gene encoding nitrate reductase: Upstream nucleotide and amino acid sequence comparisons. Gene 111:149–155

Vega JM (1976) A reduced pyridine nucleotides-diaphorase activity associated to the assimilatory nitrite reductase complex from *Neurospora crassa*. Arch Microbiol 109:237–242

Wieczorek H, Brown D, Grinstein S, Ehrenfeld J, Harvey WR (1999) Animal plasma membrane energization by proton-motive V-ATPases. Bioessays 21:637–648

Zumft WG, Dreusch A, Lochelt S, Cuypers H, Friedrich B, Schneider B (1992) Derived amino acid sequences of the *nosZ* gene (respiratory N_2O reductase) from *Alcaligenes eutrophus*, *Pseudomonas aeruginosa* and *Pseudomonas stutzeri* reveal potential copper-binding residues. Implications for the CuA site of N_2O reductase and cytochrome-c oxidase. Eur J Biochem 208:31–40