CMLS, Cell. Mol. Life Sci. 55 (1999) 450–455
1420-682X/99/030450-06 $ 1.50 + 0.20/0

CMLS Cellular and Molecular Life Sciences

Research Article

A common ancestor for a subunit in the mitochondrial proton-translocating NADH:ubiquinone oxidoreductase (complex I) and short-chain dehydrogenases/reductases

M. E. Baker[a,*], W. N. Grundy[b] and C. P. Elkan[b]

[a]Department of Medicine, 0823, University of California, San Diego, 9500 Gilman Drive, La Jolla (California 92093-0823, USA), Fax + 1 619 534 1424, e-mail: mbaker@ucsd.edu
[b]Department of Computer Science, University of California, San Diego, 9500 Gilman Drive, La Jolla (California 92093-0623, USA)

Received 26 November 1998; received after revision 12 January 1999; accepted 12 January 1999

Abstract. The proton-translocating NADH:ubiquinone oxidoreductase or complex I is located in the inner membranes of mitochondria, where it catalyzes the transfer of electrons from NADH to ubiquinone. Here we report that one of the subunits in complex I is homologous to short-chain dehydrogenases and reductases, a family of enzymes with diverse activities that include metabolizing steroids, prostaglandins and nucleotide sugars. We discovered that a subunit of complex I in human, cow, *Neurospora crassa* and *Aquifex aeolius* is homologous to nucleotide-sugar epimerases and hydroxysteroid dehydrogenases while seeking distant homologs of these enzymes with a hidden Markov model-based search of Genpept. This homology allows us to use information from the solved three-dimensional structures of nucleotide-sugar epimerases and hydroxysteroid dehydrogenases and our motif analysis of these enzymes to predict functional domains on their homologs in complex I.

Key words. NADH:ubiquinone oxidoreductase; short-chain dehydrogenases/reductases; nucleotide-sugar epimerases; hydroxysteroid dehydrogenases; evolution.

The proton-translocating NADH:ubiquinone oxidoreductase or complex I is located in the inner membranes of mitochondria [1–5]. Complex I consists of over 40 subunits that catalyze the transfer of electrons from NADH to ubiquinone. At least 20 subunits in complex I are found in a hydrophobic fraction. One of these hydrophobic subunits has been cloned and sequenced as a 40-kDa protein from *Neurospora crassa* [1] and 39-kDa proteins from cow [2] and human [3]; the function of this subunit has not been established. Here we report that this subunit in complex I is homologous to nucleotide-sugar epimerases and hydroxysteroid dehydrogenases, which are members of the short-chain dehydrogenases/reductases family (SDR) [6–9]. This family has been extensively studied because it includes enzymes that regulate the concentrations of steroids such as cortisol, estradiol and testosterone, as well as prostaglandins in humans [6–9]. As a result, there is much structure-function information about SDRs from solved three-dimensional (3D) structures of hydroxysteroid dehydrogenases [10–14] and mutagenesis studies [6, 15–17]. We use this information to predict functional domains on the ~40-kDa subunit in the hydrophobic fraction of complex I.

* Corresponding author.

Methods

Motif analysis. Motifs for a training set of 202 divergent dehydrogenases were discovered using MEME (**M**ultiple **E**xpectation-maximum for **M**otif **E**licitation), which has been described in detail elsewhere [9, 18, 19]. MEME is an artificial intelligence-based motif analysis tool that, given a set of unaligned sequences, identifies in an unbiased, automated fashion the conserved regions (i.e. motifs) that are characteristic of the dataset. Each motif is represented as a position-dependent probability matrix or log-odds matrix: each column of the matrix gives the probabilities of each residue in that position.

Database searches. The matrix representations of the motifs discovered by MEME are input into MAST (**M**otif **A**lignment and **S**earch **T**ool) and Meta-MEME [20, 21]. MAST searches databases such as Genpept and SWISSPROT with each motif independently and subsequently combines the scores. Meta-MEME incorporates the motifs into a single hidden Markov model, which is then used by HMMER [22] to conduct a Smith-Waterman search of the database. HMMER returns log-odds scores analogous to the 'bit score' returned by BLAST [23]. We also used gapped BLAST [24] to search for homologs of the $\sim$40-kDa subunit in NADH:ubiquinone oxidoreductase.

Sequence analysis. The ALIGN program, developed in Dayhoff's laboratory [25], was used to quantify the similarity between protein sequences. ALIGN calculates the best alignment between any pair of sequences using the Dayhoff scoring matrix and a penalty for breaking a sequence (gap penalty). The score for the two sequences is compared with that obtained from comparing random permutations of the two sequences. The alignment score is the number of standard deviations by which the maximum score for the real sequences exceeds the average maximum score for the random. For the analyses reported here, 10,000 random permutations were used for the statistical analysis, and the Dayhoff matrix was used with a bias of 6 and a gap penalty of 8.

Results and discussion

Sequence analysis

Hidden Markov-based database search for homologs of hydroxysteroid dehydrogenases. In the course of an evolutionary analysis of hydroxysteroid dehydrogenases and their homologs, we used MEME [9, 18, 19] to determine motifs in 202 divergent dehydrogenases. These motifs were incorporated into Meta-MEME [20, 21], a hidden Markov model database-searching tool, which was used to search Genpept for distant dehydrogenase homologs. The search assigned the 40-kDa subunit of *N. crassa* NADH:ubiquinone oxidoreductase (NUEM_NEUCR) a score of 15.5 bits. This score indicates that the sequence matches the given model $2^{15.5} = 4.6 \times 10^4$ times better than it matches a random background model. Since the Genpept database contains 280,000 sequences, of which approximately 900 are SDR homologs, any log-odds score greater than $\log_2 (280{,}000/900) = 8$ is statistically significant. Thus a score of 15.5 bits for NUEM_NEUCR indicates that this protein belongs to the SDR family.

Gapped BLAST analyses of the *N. crassa* subunit in complex I. A gapped BLAST search of Genpept of NUEM_NEUCR found two close homologs in human (NUEM_HUMAN) and cow (NUEM_BOVIN) with E values of 7×10^{-53} and 2×10^{-51}, respectively. The next closest protein is *Aquifex aoelicus* 2982870, with an E value of 9×10^{-13}. *A. aoelicus* is a thermophilic bacterium whose genome was recently sequenced [26]. *A. aoelicus* 2982870 is about 24% identical to NUEM_NEUCR and the human and bovine homologs, over a length of about 300 amino acids. *A. aoelicus* 2982870 is characterized in GenBank as an NADH ubiquinone oxidoreductase based on its BLAST score.

Further gapped BLAST analyses. NUEM_NEUCR and each of its three close homologs was used for a Gapped BLAST search of GenBank. Because these homologs are divergent, each search identified different proteins in the SDR family. The results of the searches are summarized in table 1, which shows that each complex I protein has a gapped BLAST score that shows that it is related to at least one SDR.

NUEM_HUMAN has a gapped BLAST score of 4×10^{-3} with *Nocardia* cholesterol dehydrogenase, an SDR [6, 7, 9]. NUEM_BOVIN has a gapped BLAST score of 5×10^{-5} with *S. coelicolor* e1245724, an oxidoreductase that is an SDR. NUEM_BOVIN also has a

Table 1. Sequence analysis of a subunit in the mitochondrial proton-translocating NADH:ubiquinone oxidoreductase (complex I).

Test sequence	Homologous sequence identified	Gapped BLAST score
NUEM_NEUCR	*Pseudomonas aeruginosa* WbpK	3×10^{-4}
NUEM_HUMAN	*Nocardia* cholesterol dehydrogenase	4×10^{-3}
NUEM_BOVIN	*S. coelicolor* e1245724	5×10^{-5}
NUEM_BOVIN	human 3β-hydroxysteroid dehydrogenase	10^{-3}
A. aoelicus 2982870	*A. aeolicus* 2983546	2×10^{-11}

Sequences in column 2 are either known to be members of the short-chain dehydrogenase/reductase superfamily or are described in their GenBank entry as members of this protein superfamily because they have high gapped BLAST scores with known homologs.

gapped BLAST E value of 10^{-3} with 3β-hydroxysteroid dehydrogenase. The gapped BLAST alignment shows that the two proteins are 21% identical with only five gaps. This similarity confirms Fearnley and Walker's dot matrix analysis of these two proteins, which suggested that they were homologs [27].

Gapped BLAST analysis of *A. aoelicus* 2982870 indicates that it is the closest of the four NADH:ubiquinone oxidoreductases to SDRs. *A. aoelicus* 2982870 has a gapped BLAST E value of 2×10^{-11} for *A. aeolicus* 2983546, which is listed in GenBank as a uridine diphosphate (UDP)-glucose-4-epimerase.

ALIGN analyses. Due to the unusual nature of the homology that we have found, we used another method to quantify the relationship between the subunit in NADH:ubiquinone oxidoreductases and SDRs. Pairwise comparisons of a segment of over 250 amino acids in the NADH:ubiquinone oxidoreductase subunit with corresponding segments in SDRs using the ALIGN program [25] are in agreement with the gapped BLAST results. For example, an ALIGN analysis of NUEM_BOVIN with *S. coelicolor* e1245724 yields a score of 9 SD ($p = 10^{-19}$). *A. aoelicus* 2982870 has an ALIGN score of 10.3 SD ($p < 10^{-24}$) with *A. aoelicus* UDP-glucose-4-epimerase.

Mapping of functional domains of short-chain dehydrogenases/reductases onto the hydrophobic subunit in proton-translocating NADH:ubiquinone oxidoreductase. Even distantly related proteins conserve some of their 3D structure [28, 29]. This allows information about the 3D structure of nucleotide-sugar epimerases [14] and dehydrogenases [10–12, 30] and their functional residues from mutagenesis studies [6, 7, 15–17] to predict functional domains on NUEM_HUMAN and its NADH:ubiquinone oxidoreductase homologs. With this goal in mind, we mapped the high scoring motifs from the MEME analysis of SDRs onto an alignment of the four NADH:ubiquinone oxidoreductase subunits as shown in figure 1A, along with the structures of the motifs as determined in the 3D analysis of epimerase and dehydrogenases. We also used MEME to determine six motifs for the human, *N. crassa* and *A. aoelicus* NADH:ubiquinone oxidoreductase subunits and mapped these motifs onto the alignment in figure 1B.

As seen in figure 1A, there are four dehydrogenase motifs that have strong scores with the NADH:ubiquinone oxidoreductase subunit. Motifs 1 and 6 from the dehydrogenases correspond to the $\beta\alpha\beta$ that binds the nucleotide part of NAD(P)(H) [5, 15, 29, 31–33]. Three glycines that are signatures for the turn in the $\beta\alpha\beta$ motif are shown in boxes. The Gly-Xaa-Xaa-Gly-Xaa-Xaa-Gly motif differs slightly from the Gly-Xaa-Xaa-Xaa-Gly-Xaa-Gly motif found in most SDRs, although some SDRs show this motif [6]. Such glycine-rich segments are found in the nucleotide binding domain of many oxidoreductases, including enzymes that are not descended from a common ancestor [29, 31–33]. In fact, Walker [5] proposed that this part of cow and *N. crassa* NADH:ubiquinone oxidoreductase had the $\beta\alpha\beta$ motif based on a comparison with various dehydrogenases, most of which are not SDRs. Motif 2 corresponds to β-strand D in SDRs and is part of the nucleotide cofactor binding domain [10–12]. We propose that the segment corresponding to motif 2 also is part of the nucleotide binding domain [34].

Motif 4 is β-strand F in nucleotide-sugar epimerases and hydroxysteroid dehydrogenases and other SDRs. β-strand F is just downstream of α-helix F, which contains an essential tyrosine and lysine, separated by three residues (Tyr-Xaa-Xaa-Xaa-Lys) at the catalytic site of SDRs. Twelve residues upstream from the tyrosine is a catalytically important serine [6, 8, 10–17]. We find a Tyr-Xaa-Xaa-Xaa-Lys segment in three of the four NADH:ubiquinone oxidoreductases. However, this segment is about seven residues closer to motif 4 in NADH:ubiquinone oxidoreductases than in SDRs. All four enzymes have the lysine residue that is 11 residues upstream from motif 4; in most SDRs the distance is 18 residues. All four enzymes have a serine that is 12 residues upstream from the tyrosine, which is the same distance as found in most SDRs. This serine, tyrosine and lysine in *A. aeolicus* 2982870 align with the catalytic serine, tyrosine and lysine in UDP-glucose-4-epimerase (data not shown), suggesting the identity and location of catalytically important residues on mitochondrial proton-translocating NADH:ubiquinone oxidoreductase (complex I). Tyrosine catalyzes hydride transfer; lysine is hydrogen-bonded to the 2′- and 3′-hydroxyl groups of the nicotinamide ribose; and serine is hydrogen-bonded to the catalytic tyrosine [10–17]. It will be interesting to determine the functions of the conserved serine and lysine and the partially conserved tyrosine in the NADH:ubiquinone oxidoreductase subunit.

We used MEME to identify six motifs that are characteristic of the subunit in NADH:ubiquinone oxidoreductase to identify functionally important regions in these proteins. Comparison of these motifs in figure 1B with those in figure 1A reveals that motifs 1, 2 and 3 of NADH:ubiquinone oxidoreductase correspond to the proposed nucleotide cofactor site for motifs 1, 2 and 6 of the dehydrogenases. Motifs 4, 5 and 6 of NADH:ubiquinone oxidoreductases correspond to the substrate binding domain of dehydrogenases [10–17]. However, in general, a comparison of the motifs in figures 1A and B indicates that NADH:ubiquinone oxidoreductase (complex I) has diverged substantially from most SDRs. Thus, there are likely to be unusual structures in NADH:ubiquinone oxidoreductase (complex I) compared with that determined for SDRs, thus far [6].

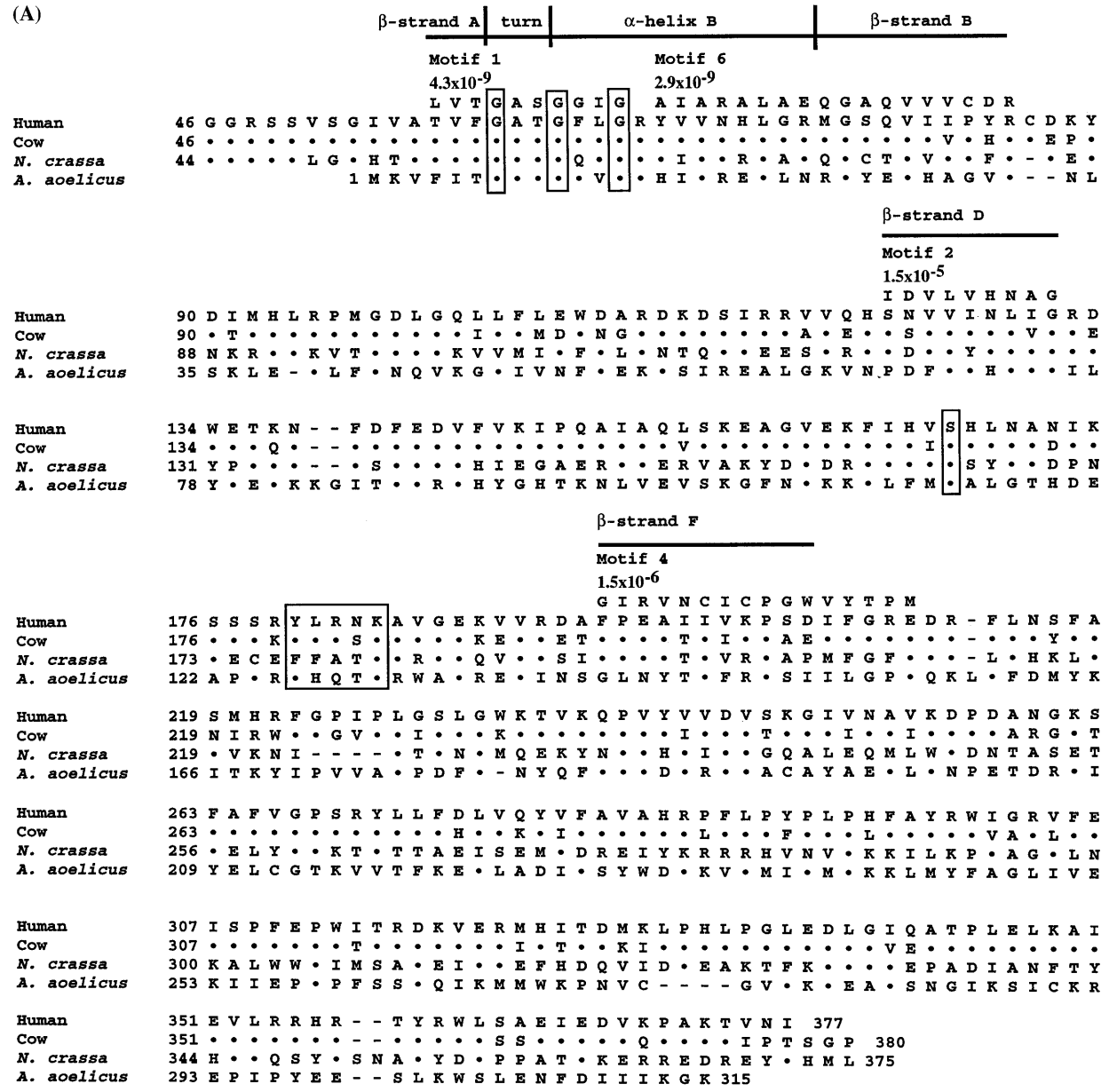

Figure 1. Mapping of MEME-generated motifs onto the ~40-kDa subunit in complex I. (*A*) Motifs determined by MEME from a training set of 202 members of the short-chain dehydrogenases/reductases family. Amino acids that are identical to those in the human subunit in complex I are denoted by a ·. Motifs were mapped onto the complex I subunit by MAST. Four motifs have significant scores and map onto the first 210 amino acids of the human sequence. The α-helices and β-strands for the motifs are based on the 3D structures of epimerases and dehydrogenases. The β-strand, turn, α-helix at the amino terminus is part of a $\beta\alpha\beta$ fold that binds the AMP moiety of NAD(P)(H) [29–32]. Conserved glycines in the $\beta\alpha\beta$ fold are boxed. The catalytically important serine and the tyrosine/lysine motif in SDRs are boxed. (*B*) Motifs determined by MEME from an analysis of the ~40-kDa subunit in complex I. MEME determined six motifs for the human, *N. crassa* and *A. aoelicus* subunit in complex I. These motifs are mapped onto an alignment of these proteins with their cow homolog. Amino acids that are identical to those in the human complex I subunit are denoted by a ·.

***A. aeolicus*, an ancient organism.** *A. aeolicus* is found in one of the most deeply branching families within the bacterial domain [26, 35], which makes its genome important from an evolutionary perspective, in addition to its value for understanding biochemical processes in an organism that grows at temperatures as high as 96 °C.

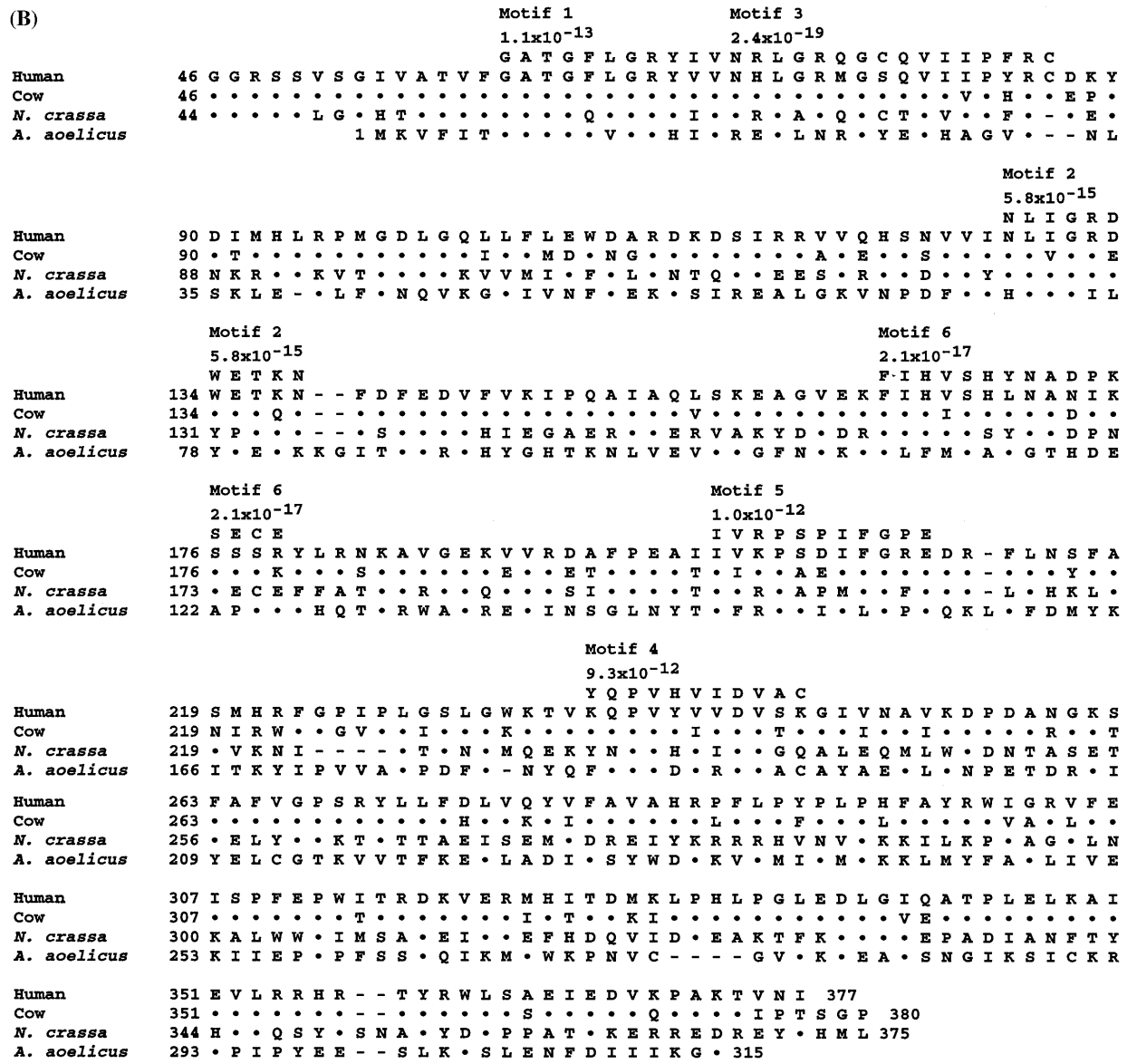

Fig. 1. Continued.

A. aoelicus contains enzymes for oxygen respiration [26], which permits it to grow in oxygen concentrations as low as 7.5 ppm. This is consistent with an ancient origin of *A. aoelicus* 298287, although it is possible that enzymes for oxygen respiration were acquired later by lateral transfer [26].

***A. aeolicus* and the origins of short-chain dehydrogenases and reductases.** The presence of UDP-glucose-4-epimerase homologs in *A. aeolicus* provides further evidence for the ancient origins of nucleotide-sugar epimerases in the SDR family. However, the ancestral enzyme for SDRs is unknown. A recent report that a protein that binds messenger RNA binding protein and is a ribonuclease is an SDR [36] suggests novel events in the origins of the SDR family.

Acknowledgments. We thank the NSF Division of Biological Instrumentation and Resources for support. W. N. Grundy is supported by a Sloan/DOE Fellowship in Computational Molecular Biology.

1 Rohlen D. A., Hoffmann J., van der Pas J. C., Nehls U., Preis D., Sackmann U. et al. (1991) Relationship between a subunit of NADH dehydrogenase (complex I) and a protein family including subunits of cytochrome reductase and a processing protease of mitochondria. FEBS Lett. **278:** 75–78

2 Fearnley I. M., Finel M., Skehel J. M. and Walker J. E. (1991) NADH:uniquinone oxidoreductase from bovine heart mitochondria. Biochem. J. **278:** 821–829

3 Baens M., Chaffanet M., Cassiman J. J., van den Berghe H. and Marynen P. (1993) Construction and evaluation of a hncDNA library of human 12p transcribed sequences derived from a somatic cell hybrid. Genomics **16:** 214–218

4 Schulte U., Fecke W., Krull C., Nehls U., Schmiede A., Schneider R. et al. (1994) In vivo dissection of the mitochondrial respiratory NADH:ubiquinone oxidoreductase (complex I). Biochim. Biophys. Acta **1187:** 121–124

5 Walker J. E. (1993) The NADH:ubiquinone oxidoreductase (complex I) or respiratory chains. Quart. Rev. Biophys. **25:** 253–324

6 Jornvall H., Persson B., Krook M., Atrian S., Gonzalez-Duarte R., Jeffrey J. et al. (1995) Short-chain dehydrogenases/reductases (SDR). Biochemistry **34:** 6003–6013

7 Baker M. E. and Blasco R. (1992) Expansion of the mammalian 3β-hydroxysteroid dehydrogenase/plant dihydroflavonol reductase superfamily to include a bacterial cholesterol dehydrogenase, a bacterial UDP-galactose-4-epimerase, and open reading frames in vaccinia virus and fish lymphocystis disease virus. FEBS Lett. **301:** 89–93

8 Krozowski Z. (1992) 11β-hydroxysteroid dehydrogenase and the short chain alcohol dehydrogenase (SCAD) superfamily. Mol. Cell. Endocrinol. **84:** C25–C31

9 Bailey T. L., Baker M. E. and Elkan C. P. (1997) An artificial intelligence approach to motif discovery in protein sequences: application to steroid dehydrogenases. J. Steroid Biochem. Molec. Biol. **62:** 29–43

10 Varughese K. I., Xuong N. H., Kiefer P. M., Matthews D. A. and Whiteley J. M. (1994) Structural and mechanistic characteristics of dihydropteridine reductase: a member of the Tyr-$(Xaa)_3$-Lys-containing family of reductases and dehydrogenases. Proc. Natl. Acad. Sci. USA **91:** 5582–5586

11 Ghosh D., Wawrzak Z., Weeks C. M., Duax W. L. and Erman M. (1994) The refined three-dimensional structure of 3α,20β-hydroxysteroid dehydrogenase and possible roles of the residues conserved in chart-chain dehydrogenases. Structure **2:** 629–640

12 Tanaka N., Nonaka T., Tanabe T., Yoshimoto T., Tsuru D. and Mitsui Y. (1996) Crystal structures of the binary and ternary complexes of 7α-hydroxysteroid dehydrogenase from *Escherichia coli*. Biochemistry **35:** 7715–7730

13 Holm L., Sander C. and Murzin A. (1994) Three sisters, different names. Nature Struct. Biol. **1:** 146–147

14 Thoden J. B., Frey P. A. and Holden H. M. (1996) Crystal structures of the oxidized and reduced forms of UDP-glactose 4-epimerase isolated from *Escherichia coli*. Biochemistry **35:** 2557–2566

15 Baker M. E. (1994) Sequence analysis of steroid- and prostaglandin-metabolizing enzymes: application to understanding catalysis. Steroids **59:** 248–258

16 Obeid J. and White P. C. (1992) Tyr-179 and lys-183 are essential for enzymatic activity of 11β-hydroxysteroid dehydrogenase. Biochem. Biophys. Res. Commun. **188:** 222–227

17 Chen Z., Jiang J. C., Lin Z. G., Lee W. R., Baker M. E. and Chang S. H. (1993) Site-specific mutagenesis of *Drosophila* alcohol dehydrogenase: evidence for involvement of tyrosine-152 and lysine-156 in catalysis. Biochemistry **32:** 3342–3346

18 Bailey T. L., and Elkan C. P. (1995) The value of prior knowledge in discovering motifs with MEME. In: Proc. Third Int. Conf. Intelligent Systems for Molec. Biology, pp. 21–29, AAAI Press, Menlo Park, CA

19 Grundy W. N., Bailey T. L. and Elkan C. P. (1996) ParaMEME: a parallel implementation and a web interface for a DNA and protein motif discovery tool. CABIOS **12:** 303–310

20 Grundy W. N., Bailey T. L., Elkan C. P. and Baker M. E. (1997) Meta-MEME: motif-based hidden Markov models of protein families. CABIOS **13:** 397–406

21 Grundy W. N., Bailey T. L., Elkan C. P. and Baker M. E. (1997) Hidden Markov model analysis of motifs in steroid dehydrogenases and their homologs. Biochem. Biophys. Res. Commun. **31:** 760–766

22 Eddy S. R. (1995) Multiple alignment using hidden Markov models. Proc. Third Int. Conf. Intelligent Systems for Molec. Biology, pp. 114–120, AAAI Press, Menlo Park, CA

23 Altschul S. F., Gish W., Miller W., Myers E. W. and Lipman D. J. (1990) Basic local alignment search tool. J. Mol. Biol. **215:** 403–410

24 Altschul S. F., Madden T. L., Schäffer A. A., Zhang J., Zhang Z., Miller W. et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402

25 Dayhoff M. O., Barker W. C. and Hunt L.T. (1983) Establishing homologies in protein sequences. Meth. Enzymol. **91:** 524–545

26 Deckert G., Warren P. V., Gaasterland T., Young W. G., Lenox A. L., Graham D. E. et al. (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. Nature **392:** 353–358

27 Fearnley I. M. and Walker J. E. (1992) Conservation of sequences of subunits of mitochondrial complex I and their relationships with other proteins. Biochim. Biophys. Acta **1140:** 105–134

28 Chothia C. and Lesk A. M. (1986) The relation between the divergence of sequence and structure in proteins. EMBO J. **5:** 823–826

29 Branden C. and Tooze J. (1991) Introduction to protein structure. Garland Publishing, New York

30 Tsigelny I. and Baker M. E. (1995) Structures important in mammalian 11β and 17β-hydroxysteroid dehydrogenases. J. Ster. Biochem. Molec. Biol. **55:** 589–600

31 Wierenga R. K., De Maeyer M. C. and Hol W. G. J. (1985) Interaction of pyrophosphate moieties with α-helixes in dinucleotide binding proteins. Biochemistry **24:** 1346–1357

32 Wierenga R. K., Terpstra P. P. and Hol W. G. J. (1986) Prediction of the occurrence of the ADP-binding $\beta\alpha\beta$-fold in proteins using an amino acid sequence fingerprint. J. Mol. Biol. **187:** 101–107

33 Scrutton N. S., Berry A. and Perham R. N. (1990) Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. Nature **343:** 38–43

34 Yamaguchi M., Belogrudov G. I. and Hatefi Y. (1998) Mitochondrial NADH-ubiquinone oxidoreductase (Complex I). Effect of substrates on the fragmentation of subunits by trypsin. J. Biol. Chem. **273:** 8094–8098

35 Pace N. R. (1997) A molecular view of microbial diversity and the biosphere. Science **276:** 734–740

36 Baker M. E., Grundy W. N. and Elkan C. S. (1998) Spinach CSP41, an mRNA-binding protein and ribonuclease, is homologous to nucleotide-sugar epimerases and hydroxysteroid dehydrogenases. Biochem. Biophys. Res. Commun. **248:** 250–254