The amount of information contained in a predictive method is an objective and comprehensive measure of its predictive power and has several advantages in evaluating the quality of a prediction method : (1) influence of database itself can be eliminated; (2) the total predictive quality and the individual predictive quality for each type of secondary structure can be assessed simultaneously; (3) contribution of all elements of the accuracy matrix are considered in a mathematically rigorous manner. It is therefore likely that a detailed information index comparison, will be useful in further improving prediction methods.

Zhi-Xin Wang

National Laboratory of Biomacromolecules, Institute of Biophysics, Academia Sinica, Beijing 100101, P.R. China

Received 6 December 1993; accepted 20 January 1994.

1. Schulz, G.E. and Schirmer, R.H. *Principles of Protein Structure*, 123 - 128, (Springer-Verlag, New York, 1979).
2. Robson, B and Garnier, J., *Introduction to Protein Engineering*, 402 - 408 (Elsevier, Amsterdam, 1986).
3. Garnier, J. and Robson, B., in *Prediction of Protein Structure and the Principles of Protein Conformation* (ed. Fasman, G.D.), 417–465 (Plenum, New York 1989).
4. Kabsch, W. and Sander, C., *FEBS Lett.*, **155**, 179–182, (1983).
5. Matthews, B. W., *Biochim. Biophys. Acta*, **405**, 442–451 (1975).
6. Rost, B. and Sander, C., *J molec Biol.*, **232**, 584–599, (1993).
7. Ash, R., *Information Theory*, 5–26 (John Wiley & Sons, New York, 1965).
8. Rost, B. and Sander, C., *Proc. natn Acad Sci* U.S.A., **90**, 7558–7562, (1993).

Three sisters, different names

Sir — The classical NAD binding fold, as seen in lactate dehydrogenase or liver alcohol dehydrogenases, consists of six strands of parallel β-pleated sheet flanked on either side by α-helices, with ~100 structurally conserved residues [1]. Here, we report a novel subclass of NAD(P)-linked oxidoreductases consisting of 3α,20β-hydroxysteroid dehydrogenase[2] (HSD), dihydropteridine reductase[3] (DHR), and UDP-galactose 4-epimerase[4] (UDP). The common core of HSD, DHR and UDP contains seven strands and five helices with more than 180 structurally equivalent residues[5] (Fig. 1). The seventh strand is linked to a classical dehydrogenase fold by a rare left-handed crossover connection. The extensive structural similarity is unexpected, as sequence identities between HSD, DHR and UDP are below 20%. Apart from common topography of the dinucleotide-binding motif, there is no close structural resemblance between this subclass and the classical dehydrogenases[2–4]. The similar folds of HSD and DHR have been noted independently by Krook *et al.*[6].

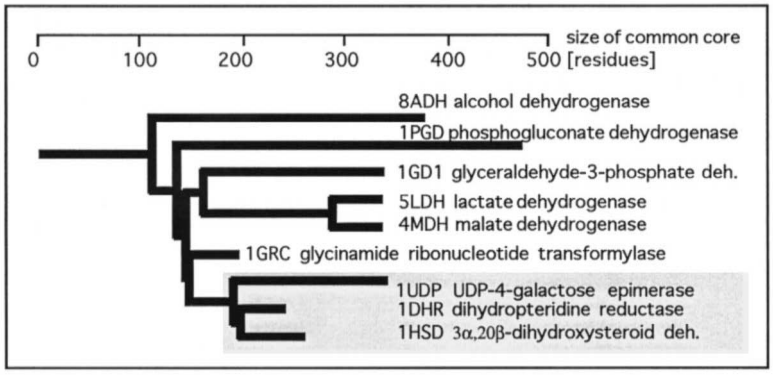

Fig. 1 Structural family tree for representative dehydrogenases. The tree was constructed by average linkage clustering of pairwise structural similarities[5]. Sequential alignment was required. The distinct subclass defined here (HSD, UDP and DHR) is shaded. UDP-galactose 4-epimerase, somewhat larger than dihydropteridine reductase and 3α,20β-hydroxysteroid dehydrogenase, has insertions in loop regions and at the carboxy-terminus that form a separate substrate binding domain. GRC is an outgroup which is topologically different from the dinucleotide binding fold (one strand has reversed direction) but strongly resembles it in overall structure.

HSD and a diverse set of bacterial and eukaryotic proteins with related sequences are classified into the insect-type alcohol dehydrogenase/ribitol dehydrogenase family[7]. Sequence relatives of UDP include sugar nucleotide epimerases, steroid dehydrogenases and flavonoid reductases of bacterial, eukaryotic and viral origin[8] while DHR is known only in man and rat. Structural alignment reveals a few strongly conserved residues common to all three families (Fig. 2). As in classical dehydrogenases, there is a glycine-rich motif in the first βαβ-unit binding the adenine part of NAD (in the crystal structures of DHR and UDP). The strictly conserved YxxxK motif on α_F is distinctive for the HSD/DHR/UDP subclass. α_F is extended at the amino terminus compared to other dehydrogenases, allowing (in DHR and UDP) the invariant tyrosine and lysine to reach over the

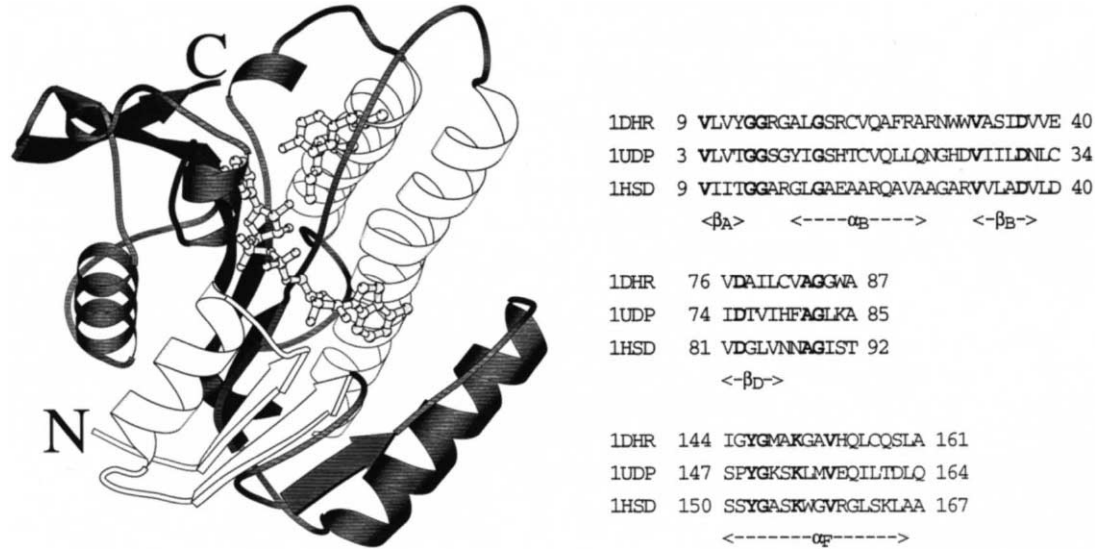

Fig. 2 Common core of HSD, UDP and DHR. *a*, Ribbon diagram[9] of dihydropteridine reductase with bound NAD (centre)[3]. The common core of HSD, DHR and UDP covers the entire fold of DHR except the C-terminal strand. *b*, Structural alignment of particularly conserved regions (bold: identical in the known structures). The conserved segments have lighter shading in the ribbon diagram. The N-terminal $\beta_A\alpha_B\beta_B$-unit (at the bottom in (*a*)) is involved in binding the adenine of NAD. In the loop after β_D (fourth strand from the bottom in (*a*)) the side chain of the conserved alanine packs against a ribose ring and the backbone NH group of the adjacent glycine interacts with a pyrophosphate unit. The side chains of the invariant tyrosine and lysine on α_F are involved in binding the nicotinamide ring (near the top in (*a*)).

edge of the β-sheet to the nicotinamide moiety of NAD. Curiously in light of the strong sequence conservation, the nicotinamide binding site as seen in DHR and UDP was presumed to be the substrate binding site in HSD[2]. Further refinement of the crystal structure of HSD may clarify this discrepancy.

The remarkable structural similarity accompanied by similarity of function and even some sequence detail strongly suggests common ancestry of HSD, DHR and UDP. The merger of three protein families into one superfamily implies structural and functional analogies useful for model building, protein engineering and understanding the evolutionary history of metabolic pathways.

Liisa Holm and Chris Sander
Protein Design Group, EMBL, D-69012, Heidelberg, Germany

Alexey Murzin
MRC Cambridge Centre for Protein Engineering and Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, UK

Received 24 November 1993; accepted 26 January 1994.

1. Ohlsson, I., Nordstrom, B. & Brändén, C.-I. *J. molec. Biol.* **89**, 339–354 (1974).
2. Ghosh, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 10064–10068 (1991).
3. Varughese, K.I., Skinner, M.M., Whiteley, J.M., Matthews, D.A. & Xuong, N.H. *Proc. natn. Acad. Sci. U.S.A.* **89**, 6080–6084 (1992).
4. Bauer, A.J., Rayment, I., Frey, P.A. & Holden, H.M. *Proteins* **12**, 372–381 (1992).
5. Holm, L. & Sander, C. *J. molec. Biol.* **233**, 123–138 (1993).
6. Krook, M., Ghosh, D., Strömberg, R., Carlquist, M. & Jörnvall, H. *Proc. natn. Acad. Sci. U.S.A.* **90**, 502–506 (1993).
7. Bairoch, A. *Nucleic Acids Res.* **19**, 2241–2245 (1991).
8. Baker, M.E. & Blasco, R. *FEBS Lett.* **301**, 89–93 (1993).
9. Kraulis, P. *J. appl. Crystallogr.* **24**, 946–950 (1991).