Volume 335, number 1, 69–72 FEBS 13293 November 1993

Variability patterns of dehydrogenases versus peptide hormones and proteases/antiproteases

Henrik Jörnvall, Bengt Persson, Hans Jörnvall*

Department of Medical Biochemistry and Biophysics, Karolinska Institutet, S-171 77 Stockholm, Sweden

Received 14 September 1993

Comparisons of human/rat species variants of 23 dehydrogenases versus 41 entries for peptide hormones and 41 for proteases/antiproteases reveal characteristic patterns. Dehydrogenases are 'constant' (2–8% residue differences between homologues) or 'variable' (12–40% differences). Peptide hormones exhibit a wider range, with many 'strictly conserved' and others spreading upscale, while proteases/antiproteases are overall shifted upscale, with few 'constant' and many 'variable'. Overall, the influence of polypeptide size and function is confirmed, while for the dehydrogenases, the data highlight the 'constant' pattern as the one with high frequency values and suggest for classic liver alcohol dehydrogenase the 'variable' pattern to correlate with emerging functions.

Evolutionary change; Protein variability; Species variation; Alcohol dehydrogenase

1. INTRODUCTION

Evolutionary changes in proteins differ by large factors, reflecting structural and functional properties of the molecules [1–3]. Within certain limits, the value for each protein is fairly constant over extensive periods, as long as overall functional properties are unaltered. Therefore, the rate of change constitutes a useful parameter in judging both protein function [1–4] and those few deviations that might indicate wide inter-species gene transfer [5].

Many dehydrogenases, however, appear to follow either of two separate evolutionary lines, being 'constant' or 'variable', with such a split also between related forms with seemingly similar properties [6]. For evaluation of which of the dehydrogenase patterns that is typical at large, and what the split may mean, comparisons with the patterns for other proteins are of interest. We have now evaluated the pattern for characterized human/rat dehydrogenases versus those for proteins with strict interactions (peptide hormones) or less strict specificities (proteases/antiproteases). Results reveal distinct patterns, emphasizing unique properties from which further conclusions can be drawn.

2. MATERIALS AND METHODS

Primary structures of pyridine-nucleotide linked dehydrogenases, proteases/antiproteases, and peptide hormones for which corresponding human and rat variants are known were extracted from the SwissProt data base [7]. Some incomplete structures from another source were also included [8]. Because of nomenclature problems, a few proteins may have been missed, and some inclusions are questionable. In the case of multi-chain native molecules, the functional activity need not always be associated with the chain(s) available, and for some of the bioactive peptides or binding proteins inclusions in the hormonal group may be unclear. However, considering the patterns observed, single missing or extra entries do not affect the conclusions. In calculations of residue differences, gap positions are counted, but in assignments of total size they are not. When the sizes of the human/rat homologues differ, the longest variant is the one listed and used in the calculation of relative differences. For native molecules with more than one type of chain, chains are listed separately when separate functions are defined (as for most hormones) but otherwise often combined (as for most proteases/antiproteases). In all cases, attempts have been made to remove signal sequences and large parts eliminated during processing, such that the entries reflect the mature forms. Especially for the hormones, differently processed forms exist, in which cases one of the common forms has been chosen for comparisons. Also, for many of the structures, conflicting reports or variants affecting single positions exist, in which case the structure used is the one from the main entry.

*Corresponding author. Fax: (46) (8) 337 462.

3. RESULTS AND DISCUSSION

Analysis of available structures for species variants of corresponding proteins in the SwissProt data base [7] identified 23 pairs of dehydrogenases characterized both in the human and rat forms (Table I), 41 pairs of entries for peptide hormones (Table II), and 41 pairs of entries for proteases/antiproteases (Table III). As stated above, single proteins may have been missed because nomenclature is not always easy to follow. Also, a few entries had to be eliminated because of multiple reports or other interpretational difficulties. The spreads of variability reveal distinct patterns as shown in Fig. 1. They have features in common, reflecting general properties of proteins, but also characteristic differences.

3.1. *Patterns for different proteins*

3.1.1. Dehydrogenases

Many dehydrogenase structures are centred in the

'constant' region, with a variability between the species in the order of about 6%, but also with a spread upscale, constituting 'variable' dehydrogenases (Fig. 1). Obviously, this overall pattern constitutes the explanation to the presence of two types of variability observed in some dehydrogenases [6]. The two types are concluded to reflect differences in functional restrictions on the structures, and the results now show that the 'constant' dehydrogenase pattern is the one with many representatives closely assembled at about the same value (Fig. 1).

Overall, the distribution of the dehydrogenases is shifted to the left in relation to those of the other two protein classes (Fig. 1). Since polypeptide chain length is one factor limiting structural variability [1], this probably reflects just the fact that dehydrogenases on the average have longer polypeptide chains than most peptide hormones and proteases/antiproteases.

3.1.2. Peptide hormones

Their pattern is clearly distributed into all variability groups. Close to one third (13 of 41) are 'strictly conserved', exhibiting no variability between the species, somewhat fewer are within the 'constant' group, while the remaining forms are widely distributed, constituting 'variable' or even 'hyper-variable' structures.

The general right-shift of the 'variable' hormones versus the 'variable' dehydrogenases probably just reflects the different sizes. However, the 'constant' hormones, at about the same variability level as the 'constant' dehydrogenases, are concluded to reflect restrictions from the functional properties, probably the importance of the receptor binding requirements. In spite of the small sizes of several peptide hormones, the multiple interactions inherent in receptor binding apparently restrict many forms, securing the selective hormone responses.

3.1.3. Proteases/antiproteases

This pattern is clearly different from those for the other two protein types, and most proteases/antiproteases characterized fall in the 'variable' group, probably reflecting the fact that the hydrolase reaction is non-complicated, not requiring coenzymes, and with often wide substrate specificities.

3.2. *Different conservation groups*

Three types of variability are apparent. A pattern of strict conservation between the human/rat protein forms is exhibited by many peptide hormones, the majority of which are fairly short (Table II), with limited elements of tertiary structure. Hence, as concluded above, it must reflect stringent needs for functional interactions, limiting the variability.

A constant pattern, at a low level of variability (2–11%), appears to be the one typical of many dehydrogenases, and also represented by the peptide hormones and a few of the proteases/antiproteases. Apparently, this is a variability level that constitutes a common background level of both large and small peptides.

Table I

Characterized structural differences between corresponding forms of human/rat dehydrogenases

Protein	Length (residues)	Gaps (positions)	Differences	
			(number)	(per cent)
Alcohol DH class I	375	1	67	18
Alcohol DH class II (fragments)	(170)	0	(37)	(22)
Alcohol DH class III	373	0	21	6
Alcohol DH class IV (fragments)	(72)	0	(11)	(15)
Aldehyde DH, cytosolic	500	0	86	17
Aldehyde DH, mitochondrial	500	0	18	4
AcylCoA DH, short-ch specific	388	0	32	8
AcylCoA DH, medium-ch specific	396	0	46	12
AcylCoA DH, long-ch specific	400	0	53	13
Glucose 6-phosphate DH	514	0	33	6
Glutamate DH	505	0	10	2
Glyceraldehyde 3-phosphate DH	334	2	21	6
3β-Hydroxy-5-ene steroid DH I	372	0	103	28
3β-Hydroxy-5-ene steroid DH II	372	1	108	29
11β-Hydroxysteroid DH	291	4	72	25
Lactate DH A	331	0	19	6
Lactate DH C	331	1	91	27
NAD(P)H DH	273	0	40	15
NADH ubiquinone DH	217	0	13	6
2-Oxoisovalerate DH, E1α subunit	401	1	19	5
3-Oxo-5α-steroid 4-DH	258	3	103	40
Pyruvate DH, E1 α subunit	361	0	7	2
Sorbitol DH	356	1	62	17

DH, dehydrogenase; ch, chain. Data from [7,8]. Alcohol DH of classes II and IV structures are still incomplete [8], as denoted by fragment designations and by values within parentheses. For glyceraldehyde 3-phosphate DH and NAD(P) DH, two different forms have been reported for the human enzymes; the one most similar to the rat enzyme structure is the one used in each case. For enzyme pairs where the subunits differ in size between the species, the length given denotes the longer subunit as indicated in the Methods section. Number of differences includes gaps, whereas per cent differences is given relative to the number of positions in the longest polypeptide within each pair. Oxoisovalerate DH and pyruvate DH are multi-chain enzymes, where the E1 components have different functions, and in the former case is a decarboxylase component, but still listed. 3-oxo-5α-steroid 4-DH is the same as steroid 5α-reductase.

A variable pattern constitutes the third type. It exhibits variabilities well above the 10% level and includes most of the proteases/antiproteases, but also many peptide hormones and dehydrogenases. Although many proteases/antiproteases frequently have a size range in between those typical of the dehydrogenases (often larger) and peptide hormones (smaller), most proteases/antiproteases fall outside the pattern typical of the latter two groups, compatible with few or variable functional

Table II

Characterized structural differences between corresponding forms of human/rat peptide hormones

Protein	Length (residues)	Gaps (positions)	Differences (number)	Differences (per cent)
ACTH	39	0	2	5
β-Lipotropin	71	0	22	31
Angiotensin	10	0	0	0
Bradykinin	9	0	0	0
Calcitonin	32	0	2	6
Cholecystokinin-33	33	0	3	9
Corticotropin RF	41	0	0	0
Erythropoietin	166	0	29	17
Gastrin	17	0	3	18
Growth hormone	191	3	66	35
Growth hormone RF	44	1	15	34
Glucagon	29	0	0	0
Inhibin, α-chain	134	1	26	19
Inhibin, β_A-chain	116	0	0	0
Insulin, A-chain	21	0	1	5
Insulin, B-chain	30	0	3	10
Insulin, C-chain	31	0	9	29
Islet amyloid polypeptide	37	0	6	16
Luteotropic hormone RH	10	0	0	0
LH, β-chain	121	6	39	32
FSH, β-chain	111	1	12	11
TSH, β-chain	112	0	10	9
Glycopeptide hormone, α-chain	96	4	26	27
Thyrotropin RH	3	0	0	0
Oxytocin	9	0	0	0
Vasopressin	9	0	0	0
Melanin concentrating hormone	19	0	0	0
Neuropeptide Y	36	0	0	0
Neurotensin-related peptide	9	0	1	11
Pancreatic hormone	36	0	8	22
Parathyroid hormone	84	0	23	27
Parathyroid hormone-related protein	141	4	18	13
Peptide YY	36	0	2	6
Prolactin	199	2	74	37
Relaxin, A-chain	24	1	11	46
Relaxin, B-chain	35	2	22	63
Secretin	27	0	3	11
Somatostatin	28	0	0	0
Thyroglobulin (fragment)	967	3	248	26
Transthyretin	127	0	22	17
Vasoactive intestinal peptide (VIP)	28	0	0	0

RF, releasing factor; RH, releasing hormone; P, protein. Two reports exist of rat insulin (parts A, B and C), rat glycopeptide hormone (α-chain), and human relaxin (A and B chains). In all cases, the structure used is the one giving highest similarity. Thyroglobin, transthyretin, islet amyloid polypeptide and some of the small peptides are not hormones, but included because of their functional roles. β-Lipotropins have an 18-residue difference N-terminally that is not included above. Reports also exist for placental lactogens but are not included since alignments are difficult, suggesting that the species variants represent non-corresponding forms. Data from [7].

Table III

Characterized structural differences between corresponding forms of human/rat proteases and antiproteases

Protein	Length (residues)	Gaps (positions)	Differences (number)	Differences (per cent)
Acrosin	274	1	66	24
Amyloid A4 (Nexin II)	753	0	22	3
α_1-Antitrypsin	394	7	124	31
Cathepsin B	260	0	45	17
Cathepsin D	348	5	52	15
Cathepsin E (fragment)	(52)	0	(11)	(21)
Cathepsin H	220	0	34	15
Cathepsin L	221	3	49	22
Chymase	227	1	89	39
Chymotrypsin B, all chains	230	0	33	14
Coagulation f IX, H-ch (fr)	227	0	35	15
Complement C3, α-chain	993	1	205	21
Complement C3, β-chain	645	3	153	24
Cystatin α	103	7	49	48
Cystatin β	98	0	21	21
Cystatin C	120	0	33	27
Cystatin S	121	3	82	68
Dipeptidyl peptidase IV (s)	731	3	114	16
Elastase 2	241	0	39	16
Furin	687	1	36	5
Glia-derived nexin (Nexin I)	379	1	57	15
Haptoglobin, α-chain	83	0	22	27
Haptoglobin, β-chain	245	0	46	19
Glandular kallikrein	237	1	97	41
Plasma kallikrein, heavy-chain	371	0	107	29
Plasma kallikrein, light-chain	248	0	50	20
Cystatin-like pep from kininogens	362	0	134	37
α_2-Macroglobulin	1451	12	389	27
Neuroendocrine convertase 1	643	1	46	7
Neuroendocrine convertase 2	529	0	17	3
Pancreatic secretory trypsin inh 1	56	0	21	38
Pepsin C (Gastricsin)	330	1	82	25
Plasminogen activator inh-1	379	0	72	19
Plasminogen activator inh-2	416	1	114	27
Renin	340	4	114	34
Thrombin	295	1	36	12
Tissue plasminogen activator	530	2	94	18
Trypsin 1	224	1	50	22
Trypsin 2	224	1	47	21
Trypsin 3	223	2	57	26
Urokinase	413	2	117	28

For the 2- and 3-chain proteases where special chain designations are not given, the entries denote the structures comprising the mature chains (plus the short activation peptides when the chains are not directly joined before cleavage). Values within parentheses indicate incomplete structures as in Table I. For glandular kallikrein, two human and three rat structures exist, for elastase 2 and pancreatic secretory trypsin inhibitor, other reports also exist. In all three cases, the pairs of most similar combinations have been used. Data from [7]; F, factor; H-ch, heavy chain; fr, fragment; s, soluble; pep, peptides; inh, inhibitor.

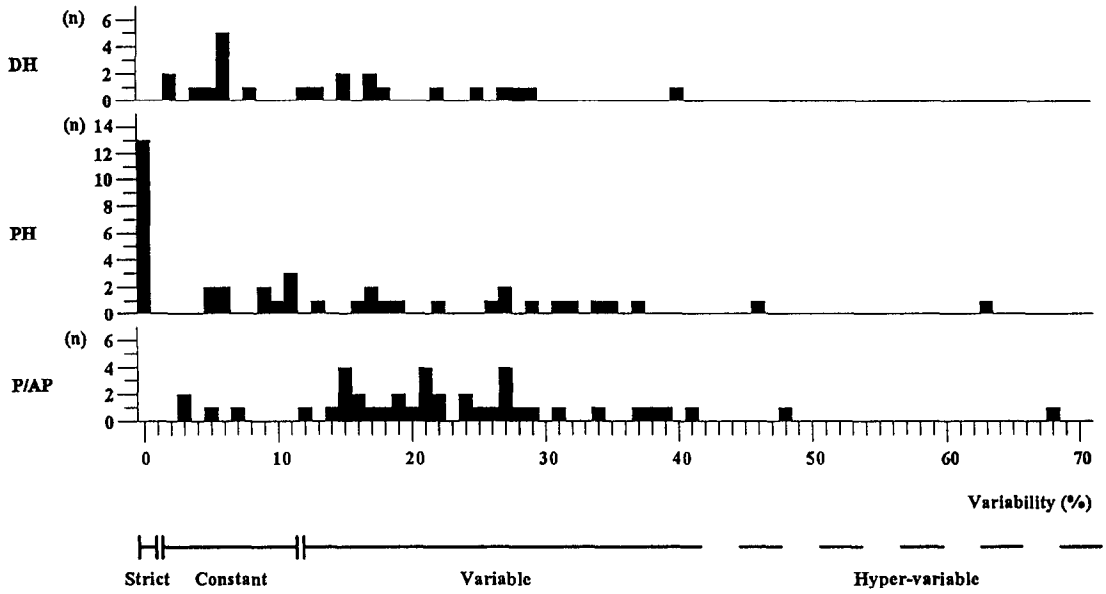

Fig. 1. Distribution of human/rat species variations between characterized protein pairs regarding dehydrogenases (DH, top), peptide hormones (PH, middle), and proteases/antiproteases (P/AP, bottom). Data from Tables I–III.

interactions in the hydrolase reaction. The variability can demonstrate a continuity to very high levels of difference, in the extreme cases above 50% between corresponding human and rat proteins (Fig. 1).

Obviously, one factor regulating all patterns is size. In addition, the correlation with functional properties is apparent. Small molecules in binding interactions are strictly conserved (Fig. 1, middle line), whereas large molecules in hydrolytic reactions can exhibit considerable variation (Fig. 1, bottom line). The three largest proteins, α_2-macroglobulin (Table III), complement C3 (Table III) and thyroglobulin (Table II), all have variable structures, with over 20% pairwise residue differences. Similarly, both large and small polypeptides (dehydrogenases and peptide hormones, respectively) demonstrate a high proportion of 'constant' proteins, apparently reflecting the 'background level' of multiple interactions.

In the case of dehydrogenases, two variability patterns, constant and variable, have been noticed [6], but at that stage it was unclear which was the common alternative. We can now see that the constant pattern appears to be the one with many representatives centred at about the same value (6%, Fig. 1), and thus the one most characteristic for that size and type of enzymatic reaction. Nevertheless, also the variable dehydrogenase type is still matched by the variabilities of the other proteins, and does therefore not indicate extraordinary circumstances, only less stringency than for the forms at the background level. It may be significant that among the dehydrogenases, class I alcohol dehydrogenase is the one variable. It is of more recent origin than class III alcohol dehydrogenase [9], with an ethanol dehydrogenase activity that has repeated origins [10] and with several characterized isozyme sets, whereas the more constant class III alcohol dehydrogenase (constituting glutathione-dependent formaldehyde dehydrogenase [11]) is of old origin and has unaltered functional properties. Apparently, a correlation with isozyme multiplicity, substrate spread, metabolic function, or time since divergence may exist.

Acknowledgements: This study was supported by grants from the Swedish Medical Research Council (03X-3532), the Swedish Alcohol Research Fund, Magnus Bergvall's Foundation and the Fund in Memory of Lars Hierta.

REFERENCES

[1] Dickerson, R.E. (1971) J. Mol. Evol. 1, 26–45.
[2] Dayhoff, M.O. (1972) Atlas of Protein Sequence and Structure, 5, National Biomedical Research Foundation, Washington DC.
[3] Doolittle, R.F. (1990) Methods Enzymol. 183.
[4] Doolittle, R.F. (1992) Prot. Sci. 1, 191–200.
[5] Doolittle, R.F. (1993) in: Methods in Protein Sequence Analysis (Imahori, K. and Sakiyama, F., Eds.) pp. 241–246, Plenum Press, New York.
[6] Yin, S.-J., Vagelopoulos, N., Wang, S.-L. and Jörnvall, H. (1991) FEBS Lett. 283, 85–88.
[7] SwissProt data bank, release 25, March 1993.
[8] Parés, X., Cederlund, E., Moreno, A., Saubi, N., Höög, J.-O. and Jörnvall, H. (1992) FEBS Lett. 303, 69–72.
[9] Danielsson, O. and Jörnvall, H. (1992) Proc. Natl. Acad. Sci. USA 89, 9247–9251.
[10] Persson, B., Bergman, T., Keung, W.M., Waldenström, U., Holmquist, B., Vallee, B.L. and Jörnvall, H. (1993) Eur. J. Biochem. 216, 49–56.
[11] Koivusalo, M., Baumann, M. and Uotila L. (1989) FEBS Lett. 257, 105–109.