Proc. Nat. Acad. Sci. USA
Vol. 70, No. 11, pp. 3230-3234, November 1973

Relatedness Among Contractile and Membrane Proteins: Evidence for Evolution from Common Ancestral Genes

(statistical methods/amino acids)

JOEL K. WELTMAN AND ROBERT M. DOWBEN*

Department of Medicine, The Miriam Hospital; and Division of Biomedical Sciences, Brown University, Providence, Rhode Island 02912

Communicated by Karl Folkers, July 9, 1973

ABSTRACT **A statistical method for quantifying the relatedness among proteins was used to perform 2926 paired comparisons of amino-acid composition among 77 contractile and membrane-associated proteins from diverse species and sources. Relatedness of amino-acid compositions correlates with homology of amino-acid sequence. A high degree of relatedness was detected among K^+-dependent membrane ATPase of *Streptococcus faecalis*, coupling factors F_1 and CF_1 from mitochondria and chloroplasts, outer fiber protein of cilia, ciliary dynein, tubulin, various actins, and myosin subfragment S-1. Heavy meromyosin and tropomyosin were related to each other but not to the first group of proteins. Differences in the degree of methylation may account for some differences in physiological function. Because of their diverse sources, the high degree of relatedness among these proteins is more compatible with evolution from common ancestral genes than with convergent evolution. Squid axon filarin, molluscan paramyosin, and bacterial flagellins appear to be unrelated either to each other or to any of the other proteins studied. Existence of persistent homologies among so many diverse proteins implies conservation of genetic information during evolution by utilization of codons for preferred amino-acid sequences in various proteins.**

Motility and membrane transport are ubiquitous processes in living organisms. These processes are mediated by specific proteins which enable part of the useful energy made available during hydrolysis of ATP (or another energy-rich compound) to be transduced into mechanical or transport work. This report concerns a study of the structural relatedness of proteins involved in motility and transport in widely different species. The results, which provide insight into the evolution of these proteins, are interesting in their own right, and also may provide clues for understanding the molecular mechanisms of energy transduction.

That some relation exists between various contractile proteins is implicit in the work of several investigators. For example, an actin-like protein has been isolated from slime mold which interacts with skeletal-muscle myosin to form a mixed actomyosin (1) and a myosin-like protein also has been extracted from slime mold which interacts with skeletal-muscle actin to form a mixed actomyosin (2). Because of similar properties of contractile proteins isolated from slime mold and from skeletal muscle as well as similar properties of skeletal-muscle actomyosin and the mixed actomyosins, it was presumed that the contractile proteins are related. Several investigators have developed the idea that actin and microtubular proteins are similar (3–5). Schnebli and Abrams (6) commented on the similar properties of membrane ATPase from *Streptococcus faecalis* and beef-heart mitochondrial ATPase.

We have undertaken to quantify the relatedness of various contractile and membrane proteins. Complete sequence data are not available for any of these proteins. However, precise values of the amino-acid composition of these proteins have been reported. The relatedness was evaluated by comparative statistical analyses of the amino-acid composition of the various proteins.

METHODS

The following classes of proteins were selected for analysis: (*1*) contractile proteins of skeletal muscle, (*2*) contractile proteins of cardiac muscle, (*3*) contractile proteins of flagella and cilia of eukaryotic cells, (*4*) bacterial flagellins, (*5*) microtubular proteins, (*6*) paramyosins of catch muscles, and (*7*) membrane-associated proteins implicated in active ion transport. In so far as available data permitted, proteins from widely different species were analyzed.

The degree of relatedness between two proteins was evaluated by statistical analyses of their overall amino-acid composition. The individual differences in mole percent content of each of 16 amino acids in two proteins were squared and summed, giving a number designated $S\Delta Q$. Methylated amino acids were added to the content of the nonmethylated amino acid from which they were derived, e.g., methyllysine and trimethyllysine were included in the lysine content. The structural and evolutionary relationships proposed in this paper must be regarded as tentative until verified by complete amino-acid sequences for the proteins.

In a previous study, the $S\Delta Q$ value was shown to correlate linearly with the degree of homology in amino-acid sequences (7). In that study, over 5000 paired comparisons were made involving over 100 different proteins. For 98% of comparison pairs, $S\Delta Q > 100$ was found for unrelated proteins. In no case was $S\Delta Q < 50$ for unrelated proteins. The theoretical maximum value for $S\Delta Q$ is that for two different amino-acid homopolymers where $S\Delta Q = 20{,}000$. In 820 comparison pairs among unrelated proteins, the median $S\Delta Q$ value was 300, and for 12% of the unrelated protein pairs compared, $S\Delta Q > 1{,}000$. The median $S\Delta Q$ values for various vertebrate hemoglobins, immunoglobulins, and cytochrome *c*'s were 80, 30, and 20, respectively.

Therefore, a value of $S\Delta Q < 50$ between two proteins is taken to indicate relatedness or homology of primary struc-

* Present address: Department of Physiology, University of Texas Southwestern Medical School, Dallas, Texas 75235.

ture. The lower the SΔQ value, the greater is the degree of relatedness. The maximum difference obtained for the same protein by the amino-acid analyses from different laboratories was SΔQ = 4; this value is the lower meaningful limit of SΔQ. SΔQ values are not metric; particularly when the molecular weights differ, they often fail to meet the requirements of a triangle ($a + b \geq c$).

RESULTS

The structural relatedness of some bacterial flagellins is shown in Table 1. Three bacterial flagellins (one from a *Salmonella* species, one from a *Bacillus*, and that from *Proteus vulgaris*) were selected to serve as index proteins for comparison. Each index protein was compared to flagellins from 17 different *Salmonella* species, 13 different *Bacillus* species, and three different amino-acid analyses of *Proteus vulgaris* flagellin. Each index protein was found to be more closely related to flagellins from bacteria of the same genus than to other flagellins. For *Proteus* flagellin, data were available only for one species. An SΔQ value of 6 was obtained, which indicates the degree of analytical variability; it compares closely to the SΔQ value of 4 or less previously found for analytical variability of amino-acid analyses of identical hemoglobin β-chains and other proteins among several laboratories (7).

As shown in Table 1, *Salmonella* flagellins and *Proteus* flagellin were more closely related to each other than flagellins from either genus were related to *Bacillus* flagellins. This finding corresponds to the observations that the metabolism, nucleotide base ratios in DNA, transformation by DNA, antigenic sites, and susceptibility to phage transduction place *Salmonella* and *Proteus* as more closely related to each other than either genus is related to *Bacillus* (8). Thus, analysis of protein relatedness by evaluation of SΔQ, as illustrated here for flagellins, may provide another basis for establishing phylogenetic relationships and for distinguishing between species.

For analysis of contractile proteins, several index proteins were selected. The analytical variability was evaluated by comparing amino-acid compositions of identical proteins from several laboratories; the results (Table 2) showed SΔQ values between 1 and 4. Comparisons between the same protein from different tissues and different species permits evaluation of polymorphism. The results (Table 2) showed constrained variability and limited polymorphism as judged by SΔQ ≤ 20, for example, between human platelet myosin and rabbit skeletal-muscle myosin. The degree of polymorphism among contractile proteins from eukaryotic cells is less than that found among bacterial flagellins. For example, the mean SΔQ among 17 *Salmonella* species compared with *Bacillus* X-1 flagellin was 73 ± 6 (Table 1).

Table 3 contains selected values of SΔQ from 2926 paired comparisons among 77 proteins. Values of SΔQ < 50 are encircled. The important relationships that emerge from these data may be summarized as follows: Squid axon filarin, molluscan catch-muscle paramyosin, and bacterial flagellins from various genetic groups are unrelated to each other and to the other proteins studied, i.e., SΔQ > 100 for two proteins from different groups. The following proteins are closely related to each other (SΔQ < 50):

TABLE 1. *Relatedness among some bacterial flagellins*

Index protein	*Salmonella* species (11–14, 43)	*Bacillus* species (15, 43)	*Proteus vulgaris* (12, 16, 43)
Salmonella paratyphi B SL 169 1,2 (11)	13 ± 8 [n = 16]	63 ± 23 [n = 13]	37 ± 7 [n = 3]
Bacillus X-1 (43)	73 ± 6 [n = 17]	33 ± 19 [n = 12]	50 ± 2 [n = 3]
Proteus vulgaris (12)	29 ± 11 [n = 17]	58 ± 17 [n = 13]	6 ± 3 [n = 2]

Mean SΔQ ± SD and number of comparisons [n] performed against index protein. Numbers in parentheses refer to the references from which amino-acid composition data were taken.

TABLE 2. *Variability and polymorphism among contractile proteins*

	Compared with	
Rabbit skeletal actin (25)	Rabbit skeletal actin 4 ± 0 [n = 2]	Other or actins* 18 ± 10 [n = 3]
Rabbit skeletal myosin (20)	Rabbit skeletal myosin (9, 23) 1 ± 1 [n = 2]	Other or myosins† 16 ± 10 [n = 4]
Rabbit skeletal tropomyosin (24)	Other tropomyosins‡ 4 ± 2 [n = 5]	
Rabbit cardiac myosin (23)	Rabbit skeletal myosin (9, 20, 23) 2 ± 1 [n = 3]	
Scallop (*Pecten*) outer fiber (17)	Sea urchin (*Strongylocentrotus*) outer fiber (17) 2.0 [n = 1]	
Clam (*Venus*) paramyosin (28)	Earthworm (*Lumbricus*) paramyosin (28) 26.4 [n = 1]	
Human platelet myosin (22)	Rabbit skeletal myosin (23) 20 [n = 1]	
Squid (*Dosidicus*) axon tubulin (29)	Protozoa (*Tetrahymena*) tubulin (29) 6.1 [n = 1]	

Mean SΔQ ± SD and number of comparisons performed [n]. Numbers in parentheses refer to references from which amino-acid composition data were taken.

* Scallop (*Pecten* (17), squid (18), and slime mold (*Plasmodium*) (19) actins.

† Lobster (*Homarus*) (20), slime mold (*Plasmodium*) (21), human platelet (22), and rabbit cardiac (23) myosins.

‡ Tropomyosins from ox, sheep-skeletal, and cardiac muscle, and from fowl smooth muscle (24).

TABLE 3. SΔQ *comparisons among contractile and membrane proteins*

	Squid (*Dosidicus*) axon filarin	Squid (*Dosidicus*) axon tubulin	*Streptococcus faecalis* ATPase	Beef heart coupling factor F_1	Chloroplast coupling factor CF_1	*Salmonella* flagellin	*Proteus* flagellin	*Bacillus* flagellin	*Spirillum* flagellin	Scallop (*Pecten*) outer fiber
Filarin (30)	0	85	95	134	147	200	141	110	190	75
Tubulin (29)	85	0	(19)	(36)	51	163	127	105	134	(10)
ATPase (6)	95	(19)	0	(14)	(23)	123	95	79	108	(14)
F_1 (31)	134	(36)	(14)	0	(19)	140	131	113	102	(34)
CF_1 (32)	147	51	(23)	(19)	0	176	150	128	118	(47)
Salmonella flagellin (11)	200	163	123	140	176	0	(43)	67	(34)	149
Proteus flagellin (12)	141	127	95	131	150	(43)	0	51	63	116
Bacillus flagellin (43)	110	105	79	113	128	67	51	0	73	106
Spirillum flagellin (15)	190	138	108	102	118	(34)	63	73	0	139
Outer fiber (17)	75	(10)	(14)	(34)	(47)	149	116	106	139	0
Tetrahymena dynein (*)	83	58	53	92	116	224	156	133	241	(41)
Rabbit actin (25)	115	(21)	(20)	(29)	(45)	155	142	120	134	(19)
Rabbit myosin (20)	58	98	79	105	114	222	188	110	219	77
Clam paramyosin (28)	164	263	239	268	261	295	275	169	287	234
HMM (9)	61	62	(47)	68	88	196	167	111	197	(45)
S-1 (32)	96	(41)	(30)	(46)	75	184	157	127	185	(28)
LMM (9)	172	308	274	315	294	449	395	260	438	271
Tropomyosin (24)	211	384	344	378	375	462	439	290	468	352
Troponin (33)	79	149	143	158	189	321	280	179	309	127

Values of SΔQ less than 50 are encircled. Numbers in parentheses give references from which amino-acid compositions were taken. *HMM* and *LMM*, heavy and light meromyosin, respectively; *S-1*, fragment of myosin.
* Gibbons, I. R., unpublished.

Group I—*Streptococcus faecalis* membrane ATPase, mitochondrial F_1 coupling factor, ciliary outer fiber protein, tubulin, actin, S-1 fragment of myosin, and Spinach chloroplast CF_1 coupling factor. Dynein is related to the S-1 myosin fragment but not to the other proteins comprising group I. All proteins in group I may be regarded as possessing an ATPase activity of some type.

Group II—Light meromyosin and tropomyosin. It is noteworthy that the α-helix content of these two proteins approaches 100% (9, 10), and yet they are not related to paramyosin, which also is almost 100% α-helical (10).

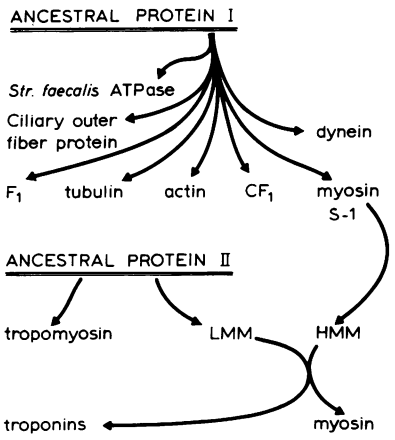

FIG. 1. A schematic diagram of the possible evolutionary development of contractile and membrane proteins based on evaluation of their relatedness by statistical analyses of amino-acid composition. The proteins are ordered in the diagram according to the relative degree of relatedness. *LMM* and *HMM*, light and heavy meromyosin, respectively.

DISCUSSION

Although statistical analysis of amino-acid sequences is the most definitive means of assessing relatedness between proteins (34), it frequently is necessary to use other methods of evaluation because amino-acid sequence data are not available. The statistical method based upon differences in total amino-acid composition used in this report has been shown empirically to correlate very closely with relatedness assessed from amino-acid sequences in three groups of proteins: hemoglobins, light chain immunoglobulins, and cytochromes (7). For proteins within these groups, a highly significant correlation was shown between SΔQ and amino-acid differences. Furthermore, more than 100 randomly selected other proteins did not show relatedness to the above groups or to each other. The correlation between SΔQ and amino-acid sequence must result from nonrandomness of amino-acid distribution in proteins; SΔQ should be invariant of permutations of amino acids, and would not *a priori* be expected to imply order of sequences. The SΔQ method has been used by Marchalonis (35) for evolutionary studies of immunoglobulin μ-chains, by Prager and Wilson (36) for studies of lysozyme phylogeny, by Robert *et al.* (37) for comparative studies of membrane glycoproteins and transplantation antigens, and by Zagalsky (38) to assess the relatedness of carotenoid-containing lipoglycoproteins isolated from ovaries and eggs of various invertebrates.

The SΔQ method does not permit a distinction between homologies of structure resulting from descent from a common ancestor, and analogies of structure resulting from con-

TABLE 3. *(cont'd)*

	Dynein	Actin	Myosin	Paramyosin	HMM	S-1	LMM	Tropomyosin	Troponin
(Filarin)	83	115	58	164	61	96	172	211	80
(Tubulin)	58	(21)	98	263	62	(41)	308	384	149
(ATPase)	53	(20)	79	239	(47)	(30)	274	344	143
(F_1)	92	(29)	105	268	68	(46)	312	378	158
(CF_1)	116	(45)	114	261	88	75	294	375	189
(*Salmonella* flagellin)	224	155	222	295	196	184	449	462	321
(*Proteus* flagellin)	156	142	188	275	167	157	395	439	280
(*Bacillus* flagellin)	133	120	110	169	111	127	260	290	179
(*Spirillum* flagellin)	241	134	219	287	197	185	438	468	309
(Outer fiber)	(41)	(19)	77	234	(45)	(28)	271	352	127
(Dynein)	0	66	60	245	(35)	(24)	231	298	104
(Actin)	66	0	121	307	75	(39)	354	431	180
(Myosin)	60	121	0	107	(12)	56	76	107	(27)
(Paramyosin)	245	307	107	0	173	257	83	99	120
(HMM)	(35)	75	(12)	173	0	(18)	146	183	(45)
(S-1)	(24)	(39)	56	257	(18)	0	253	304	102
(LMM)	231	354	76	83	146	253	0	(30)	74
(Tropomyosin)	298	431	107	99	183	304	(30)	0	107
(Troponin)	104	180	(27)	120	(45)	102	74	107	0

vergence to common properties because of similar function and selective pressures. Where genetic data are available, low SΔQ values have been associated with common ancestry. Furthermore, proteins from diverse sources with different function, such as the contractile and membrane proteins evaluated in this study, are less likely to converge to similar composition than to have evolved from a common ancestral source. The existence of persistent homologies among so many diverse proteins is indicative of conservation of genetic information during evolution while the specificity of function was altered.

The sets of proteins organized into groups I and II may most simply be related to each other according to the scheme depicted in Fig. 1, which implies that many of the proteins arose from two ancestral genes, one for the group I proteins (those possessing some type of interaction with ATP) and one for the group II proteins (possessing almost 100% α-helical structure). For muscle actin, a group I protein, it should be remembered that 1 mole of ATP is hydrolyzed to ADP when a mole of actin monomer polymerizes to fibrous actin. During sonication, muscle actin has sustained ATPase activity (39), and the Mg-polymer of slime mold actin has sustained ATPase activity (40) without sonication.

Owing to the fact that the molecular weight of the light meromyosin portion of myosin (9) is roughly three times as large as tropomyosin (41), it is not unlikely that the light meromyosin portion arose as a result of gene duplication. The entire large myosin subunit, then, may have been formed by gene fusion of light meromyosin and S-1 subfragment, with some interspersed peptide region. It may be that the middle portion of the complete myosin gene was then used as a structural basis for the troponin-like regulatory proteins. This view is supported by the SΔQ value of 27 for troponin compared with whole myosin, a lower value of SΔQ than those obtained when troponin is compared with light meromyosin, heavy meromyosin, or S-1 subfragment.

For rabbits, a comparison of skeletal-muscle myosin and cardiac myosin yields SΔQ = 2, a value within the limits of amino-acid analytical variability (Table 2). Thus, the basic backbone of amino acids comprising these proteins may be regarded as being virtually identical. It is well known, however, that the contractile proteins of skeletal and cardiac muscle differ in their percentage of methylated amino-acid residues (42). The differences in the degree of methylation may account for the differences in physiological properties of these contractile proteins, even though the backbone structure may be closely homologous. This represents another instance of the conservation and reutilization of genetic information.

In view of the striking relatedness of the proteins comprising group I despite their diverse origins, it would not be surprising if some common denominator were found in the molecular mechanisms by which these proteins enable energy transductions to take place.

Supported by grants from the National Science Foundation (GB-30505), the National Institute of Arthritis and Metabolic Diseases (AM-13438) and The Miriam Hospital Research Fund.

1. Hatano, S. & Oosawa, F. (1966) *J. Cell. Physiol.* **68**, 197–202.
2. Adelman, M. R. & Taylor, E. W. (1969) *Biochemistry* **8**, 4976–4988.
3. Renaud, F. L., Rowe, A. J. & Gibbons, I. R. (1968) *J. Cell Biol.* **36**, 79–90.
4. Mazia, D. & Ruby, A. (1968) *Proc. Nat. Acad. Sci. USA* **61**, 1005–1012.
5. Puszkin, E., Puszkin, S. & Aledort, L. M. (1971) *J. Biol. Chem.* **246**, 271–276.
6. Schnebli, H. P. & Abrams, A. (1970) *J. Biol. Chem.* **245**, 1115–1121.
7. Marchalonis, J. J. & Weltman, J. K. (1971) *Comp. Biochem. Physiol.* **38**, 609–625.

8. Jones, D. & Sneath, P. H. A. (1970) *Bacteriol. Rev.* **34,** 40–81.
9. Lowey, S. & Cohen, C. (1962) *J. Mol. Biol.* **4,** 293–308.
10. Cohen, C. & Szent-Gyorgyi, A. G. (1957) *J. Amer. Chem. Soc.* **79,** 248–252.
11. McDonough, M. W. (1965) *J. Mol. Biol.* **12,** 342–355.
12. Chang, J. Y., Brown, D. M. & Glazer, A. N. (1969) *J. Biol. Chem.* **244,** 5196–5200.
13. Ada, G. L., Nossal, G. J. V., Pye, J. & Abbot, A. (1964) *Aust. J. Exp. Biol. Med. Sci.* **42,** 267–282.
14. Parish, C. R. & Ada, G. L. (1969) *Biochem. J.* **113,** 489–499.
15. Martinez, R. J., Brown, D. M. & Glazer, A. N. (1967) *J. Mol. Biol.* **28,** 45–51.
16. Rinker, J. N. (1957) Ph.D. Thesis, Purdue University.
17. Stephens, R. E. & Linck, R. W. (1969) *J. Mol. Biol.* **40,** 497–501.
18. Taylor, E. W. (1968) *Neurosci. Res. Program Bull.* **6,** 156–161.
19. Hatano, S. & Oosawa, F. (1966) *Biochim. Biophys. Acta* **127,** 488–498.
20. Kominz, D. R. (1966) in *Phylogeny of Immunity,* eds. Smith, R. T., Miescher, P. A. & Good, R. A. (University of Florida Press, Gainsville, Fla.), pp. 49–76.
21. Hatano, S. & Ohnuma, J. (1970) *Biochim. Biophys. Acta* **205,** 110–120.
22. Booyse, F. M., Hoveke, T. P., Zschocke, D. & Rafelson, M. E. (1971) *J. Biol. Chem.* **246,** 4291–4297.
23. Barany, M., Gaetjens, E., Barany, K. & Karp, E. (1964) *Arch. Biochem. Biophys.* **106,** 280–293.
24. Woods, E. F. (1969) *Biochemistry* **8,** 4336–4344.
25. Adelstein, R. S. & Kuehl, W. M. (1970) *Biochemistry* **9,** 1355–1364.
26. Laki, K., Maruyama, K. & Kominz, D. R. (1962) *Arch. Biochem. Biophys.* **98,** 323–331.
27. Carsten, M. E. & Katz, A. M. (1964) *Biochim. Biophys. Acta* **90,** 534–541.
28. Laki, K. (1971) in *Contractile Proteins and Muscle,* ed. Laki, K. (Dekker, Inc., New York), pp. 273–306.
29. Davison, P. F. & Huneeus, F. C. (1970) *J. Mol. Biol.* **52,** 429–439.
30. Huneeus, F. C. & Davison, P. F. (1970) *J. Mol. Biol.* **52,** 415–428.
31. Racker, E. (1970) "Membranes of mitochondria and chloroplasts," in *Amer. Chem. Soc. Monograph No. 165* ed. Racker, E. (Van Nostrand-Reinhold, Princeton, N.J.), pp.127–171.
32. Lowey, S., Slayter, H. S., Weeds, A. G. & Baker, H. (1969) *J. Mol. Biol.* **42,** 1–29.
33. Ebashi, S. & Endo, M. (1968) *Progr. Biophys. Mol. Biol.* **18,** 123–184.
34. Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Silver Spring, Md.).
35. Marchalonis, J. J. (1972) *Nature New Biol.* **236,** 84–86.
36. Prager, E. M. & Wilson, A. C. (1971) *J. Biol. Chem.* **246,** 5978–5989.
37. Robert, L., Sayolle, J., Derovette, F. & Zabriskie, J. (1972) *Transplant. Proc.* **4,** 415–418.
38. Zagalsky, P. S. (1972) *Comp. Biochem. Physiol.* **41,** 385, 397–395.
39. Asakura, S. & Oosawa, F. (1960) *Arch. Biochem. Biophys.* **87,** 273–280.
40. Totsuka, T. & Hatano, S. (1970) *Biochim. Biophys. Acta* **223,** 189–197.
41. Spudich, J. A. & Watt, S. (1971) *J. Biol. Chem.* **246,** 4866–4871.
42. Trayer, J. P., Harris, C. I. & Perry, S. V. (1968) *Nature* **217,** 452–453.
43. Smith, R. W. & Koffler, H. (1971) *Adv. Microb. Physiol.* **6,** 241.