in trace metals (*14*). *Spartina* litter also releases fulvic acids, or precursors, during decomposition, and such organic decomposition products could chelate dissolved trace metals which collect at the surface microlayer and become available for scavenging by the litter of the marsh (*14*). Thus, *Spartina* litter appears to both sorb surface microlayer components and release organic materials which comprise an integral part of the microlayer. Such organic materials could provide natural soluble chelators for trace metals and account in part for net export as dissolved components.

Earlier surveys on salt marshes to the south (*4*) suggested less of a range of trace metal tidal fluxes. At either locale, there is probably little long-term net export or import of most trace metals, except possibly iron. In the Delaware salt marsh, iron appears to undergo large (100×) dissolved export during periods of the summer; this export is due to acid-producing reactions of sulfide oxidation at the salt marsh sediment surface (*15*). Tidal heights in this salt marsh are moderate (1.5 m), and often inundations of large areas of the salt marsh surface are only monthly. Thus the export of trace metals entrained in the marsh by the action of the surface microlayer or arising from oxidation at the sediment surface may be more periodic. Where there is greater tidal height and thus complete inundation more often, exports of surface-entrained materials might be more regular. For any salt marsh, however, reactions at the tidally regulated salt marsh surface seem critical to the flux and cycling of trace metals and other elements.

ROBERT E. PELLENBARG
THOMAS M. CHURCH
College of Marine Studies, University of Delaware, Newark 19711

References and Notes

1. V. L. Chapman, *Salt Marsh and Salt Deserts of the World* (Interscience, New York, 1960).
2. E. P. Odum and A. A. de la Cruz, *AIBS Bull.* **13**, 39 (1963); G. Tyler, *Oikos* **22**, 265 (1971).
3. R. J. Reimold, *Hydrobiologia* **36**, 361 (1970).
4. J. L. Settlemeyre and L. R. Gardner, in *Marine Chemistry in the Coastal Environment*, T. M. Church, Ed. (Symposium Series, American Chemical Society, Washington, D.C., 1975), vol. 18, p. 152.
5. A. A. de la Cruz and W. E. Poe, *Limnol. Oceanogr.* **20**, 124 (1975); L. P. Gardner, *ibid.*, p. 81.
6. S. R. Piotrowciz, B. J. Ray, G. L. Hoffman, R. A. Duce, *J. Geophys. Res.* **77**, 5243 (1972).
7. K. H. Szekielda, *et al.*, *ibid.*, p. 5278. D. F. Polis, *ibid.*, p. 5278.
8. G. T. Wallace and R. A. Duce, *Mar. Chem.* **3**, 157 (1975).
9. F. MacIntyre, *Sci. Am.* **223** (No. 5), 104 (1974).
10. R. E. Pellenbarg and T. M. Church, *Anal. Chim. Acta* **97**, 81 (1978).
11. We calculated the trace metal budgets for the Canary Creek salt marsh, using the mass of water fluxing from the marsh as integrated between periodic cross-sectional current measurements. We calculated the flux of dissolved metal by integrating the concentration in grams of metal per gram of liquid water; the seston metal flux was calculated from the grams of metal per gram of dry seston multiplied by integrated mass flux of seston. We calculated the amount of surface microlayer flux by multiplying the grams of metal per freeze-dried mass of microlayer material occupying a density of 1.6 g/cm^3 by the surface microlayer volume (assumed to have a thickness of 1.5 μm) by the area of the marsh (1.93 km^2). During maximum inundations, the water in the floodplain is assumed to suspend the same quantity of fine-grained, easily mobilized seston as that which exits the terminus of the marsh at the time of collection. Also, the surface microlayer as collected is assumed to occupy a uniform thickness of 1.5 μm over the marsh. The mass of microlayer used in the flux calculation is that analyzed after freeze-drying rather than the mass collected, which contains some entrained water. The actual microlayer is probably a hydrophobic film which contains little of this water and is composed primarily of large organic molecules and particulates whose mass is little affected by the process of freeze-drying.
12. The differences between the fluxes of the three metal components (flood minus ebb) yield positive (import) or negative (export) residuals. However, these residuals are probably significant only in sign, relative to the components sampled (surface microlayer, dissolved components, and seston). These signs were consistent for all three seasonal tides sampled. Whether they are representative on a longer term basis or for salt marshes in general cannot be determined until more data have been acquired.
13. J. L. Gallagher, *Limnol. Oceanogr.* **20**, 120 (1975).
14. R. E. Pellenbarg, *Estuarine Coastal Mar. Sci.* **6**, 187 (1978).
15. C. Lord and T. M. Church, in preparation.
16. Supported by an NSF grant from the Ocean Science Division (DES-74-21512).

19 August 1977; revised 15 June 1978

β-Galactosidase and Selective Neutrality

Abstract. *Three hypotheses to explain the amino acid composition of proteins are inconsistent* ($P \simeq 10^{-9}$) *with the experimental data for β-galactosidase from* Escherichia coli. *The exceptional length of this protein, 1021 residues, permits rigorous tests of these hypotheses without complication from statistical artifacts. Either this protein is not at compositional equilibrium, which is unlikely from knowledge about other proteins, or the evolution of this protein and its coding gene have not been selectively neutral. However, the composition of approximately 60 percent of the molecule is consistent with either a selectively neutral or nonneutral evolutionary process.*

Although the amino acid composition is a readily measured characteristic of protein structure, the factors that contribute toward determining this composition are not well understood. Various hypotheses have been put forward to explain the experimental compositions. If the replacement of one amino acid by another is a neutral process with respect to Darwinian selection, then at compositional equilibrium the amino acids would be, within statistical scatter, equally abundant in proteins at a frequency of 1/20. In contrast, if selection occurs at the gene level and is a neutral process, the four nucleotides in the structural gene would be approximately equally abundant at a frequency of 1/4 each, that is, in a ratio of 1:1:1:1, at each of the three coding triplet positions. In this case, the expected amino acid compositions would be very close to that in the genetic code table, exclusion of the chain-terminating codons perturbing this result only slightly. Finally, independent of the extent to which evolution has been selectively neutral, the observed amino acid composition in individual proteins could be considered a sampling from the average natural abundances of amino acids in proteins. For brevity, these three explanations will be designated the amino acid–random, genetic code–random, and natural abundance–random hypotheses, respectively.

In the aggregate, the proteins that have been sequenced so far do not support any of these three hypotheses (*1–3*). A more sensitive examination of each would be possible if the amino acid composition of individual proteins could be tested against them. In some of the referenced studies such a sensitive test could not be made because the distribution of the test statistic was not known with certainty, or because the number of amino acid residues in particular proteins is so small as to reduce the force of the inferences drawn. The latter consideration is particularly relevant with respect to determining whether a particular amino acid type is or is not present in excess or deficit of theoretical expectation. The most satisfactory way to avoid these problems is to test each hypothesis against a protein of very long length so that (i) the effect of short finite length is statistically negligible; (ii) the assumptions under which the distribution of the χ^2 test statistic are valid, are satisfied; and (iii) in testing a particular amino acid against these hypotheses, the binomial distribution (see below) is well approximated by the normal distribution.

The amino acid sequence of β-galactosidase from *Escherichia coli* reported by Fowler and Zabin (*4*) provides an opportunity to examine the several hypotheses about the compositional structure of proteins. The unusual length of this protein, 1021 residues, permits the testing of each, statistically free of the sampling error introduced by small sample size. Also, as there are 20 amino acid types, each is represented in the sequence by the order of 50 residues, which is a sufficiently large number in itself to make meaningful statements about each amino

0036-8075/79/0309-1012$00.50/0

acid type. Conditions (i), (ii), and (iii) mentioned above are satisfied. In this report, I give the results of testing the amino acid–, genetic code–, and natural abundance–random explanations of amino acid composition against the experimental composition of β-galactosidase from *E. coli*.

If the amino acids are equally frequent on the average, the expected number of each is 1021/20 or 51.05. The experimentally observed frequencies are in the second column of Table 1 ($\chi^2 = 159.24$). With 19 degrees of freedom, the probability of observing the experimental results, were the hypothesis true, is $P = 2.2 \times 10^{-9}$. The amino acid–random hypothesis is rejected.

If the amino acids are present at their genetic code table frequencies, the expected number of each is $n_i/61$, where n_i is the codon multiplicity for amino acid i. These expected compositions are in the third column of Table 1. They are not in agreement with the experimental amino acid composition: $\chi^2 = 158.10$, d.f. = 19, $P = 2.3 \times 10^{-9}$. The hypothesis that on the average the nucleotide triplets that code for the amino acids are used equally frequently is rejected, in agreement with experimental nucleic acid sequence data (*5*) for genes which code for other proteins in mammals.

Perhaps the most reasonable of the three hypotheses is the last. These natural abundances (*6*) at least have the virtue that they are the direct averages of experimentally observed, rather than theoretical or assumed, amino acid compositions. Nevertheless, the last column in Table 1 demonstrates conclusively ($\chi^2 = 157.0$, d.f. = 19, $P = 2.8 \times 10^{-9}$) that the amino acid composition of β-galactosidase from *E. coli* is not simply a sampling from the compositional pool given by these natural abundances.

A simpler, but less statistically rigorous, way of looking at the data is to calculate the average absolute (that is, disregarding algebraic sign) percentage of deviation of the expected compositions from that observed. The results for the three hypotheses discussed above are, in the order given, 48, 32, and 37 percent. A still simpler manner to view the data is to count the minimum number of steps (amino acid replacements) per 100 residues

$$Q = \sum_{i=1}^{20} |n_{io} - n_{ie}| \times \frac{100}{T}$$

it takes to get from the expected composition to that actually observed, where n_{io} and n_{ie} are the observed and expected number of residues of amino acid i and T is the protein length in residues. The results are 34, 28, and 27 steps per 100 codons, respectively, for the amino acid–, genetic code–, and natural abundance–random hypotheses. β-Galactosidase is quite dissimilar to any of the posited compositions. These commonsense calculations are in full agreement with the rigorous analyses of the preceding paragraphs.

Table 1. Comparison of observed and expected amino acid composition of β-galactosidase under two hypotheses. (i) Amino acid frequency is proportional to the codon frequencies in the genetic code table. (ii) Amino acid frequency is proportional to the natural abundances of the amino acids in proteins. There are 1021 residues in β-galactosidase.

Amino acid	Observed frequency	Expected frequency*	
		Genetic code	Natural abundance
Alanine	76	66.95	88.70
Arginine	66	100.42−	40.17+
Asparagine	46	33.48	50.21
Aspartic acid	65	33.48+	60.26
Cysteine	16	33.48−	25.11
Glutamine	60	33.48+	43.52+
Glutamic acid	60	33.48+	58.58
Glycine	70	66.95	78.66
Histidine	34	33.48	21.76+
Isoleucine	39	50.21	56.91−
Leucine	96	100.42	75.31+
Lysine	20	33.48−	70.30−
Methionine	23	16.73	16.74
Phenylalanine	38	33.48	38.50
Proline	63	66.95	38.50+
Serine	60	100.42−	70.30
Threonine	56	66.95	60.26
Tryptophan	39	16.73+	13.39+
Tyrosine	31	33.48	43.52
Valine	63	66.95	70.30
Total	1021	1021	1021
	χ^2	158.10	156.97
	Degrees of freedom	19	19
	Probability	2.28×10^{-9}	2.84×10^{-9}

*A plus or minus following an expected frequency indicates that amino acid is in excess (+) or deficit (−) of expectation at the 1 percent level of statistical significance.

The significant deviations of individual amino acid types from their expected values can be found by using the binomial statistic derived by Laird and Holmquist (*7*). As their published tables of critical values only go up to proteins 600 residues in length, those tables can be extended by noting that at the 1 percent level of statistical significance, for proteins 600 or more residues in length, the binomial statistic can be excellently approximated by the normal distribution with mean Te_i and variance $Te_i(1 - e_i)$, where e_i is the expected proportion of amino acid i. Under the amino acid–random hypothesis, at the 1 percent level of significance the observed values should lie between 51.05 ± 16.20. Alanine, glycine, and leucine are in excess of expectation and cysteine, histidine, lysine, methionine, and tyrosine are significantly less than expected. Under the genetic code–random hypothesis aspartic acid, glutamic acid, glutamine, and tryptophan are in excess and arginine, cysteine, lysine, and serine are in deficit (Table 1). Under the abundance-random hypothesis arginine, glutamine, histidine, lysine, phenylalanine, and tryptophan are in excess, and isoleucine and lysine in deficit. Thus assignable selective effects are concentrated among eight amino acid types representing 386 (genetic code random) or 417 (natural abundance random) of the total of 1021 residues. These amino acids account for 90 percent of the total chi-square value. The chi-square value of 14.15 (15.35) for the remainder of the molecule does not differ from expectation. The evolutionary process that formed this part of the molecule may or may not have been selectively neutral.

The above results indicate that all three of the explanations put forward in an attempt to explain the amino acid composition of proteins are equally bad, the probability of any of them being correct being little better than one in a billion. Either the structure of β-galactosidase is not at compositional equilibrium or the usual explanations are wrong.

In all known protein families, each individual member of a given family has an almost identical biological function with respect to any other member of that family: cytochrome c serves the same electron transport function in man as in fungi. As structure determines function, this functional equilibrium must reflect a structural equilibrium, at least with respect to amino acid composition, and for some aspects of sequence as well (in-

varient residues, for example). The fact that all cytochrome c's have an isoelectric point near *p*H 9 is one manifestation of this compositional equilibrium, as is the fact that, for a given amino acid, the number of residues of that amino acid in any cytochrome c does not vary wildly from the mean value for the family. Specific sequence differences between proteins of the same family appear to be functional fine tuning resulting from the process of natural selection with respect to a given species (*8*). Similar statements hold for the hemoglobin chains, parvalbumins, lysozymes, and other protein families. There is no reason to believe that β-galactosidase is exempt from these general considerations (the preponderance of shorter proteins in previous compilations of protein structure is due to the fact that shorter proteins are easier to sequence).

The conclusion is that the observed amino acid composition of β-galactosidase from *E. coli* is not the sole result of a selectively neutral evolutionary process with respect to (i) base replacement at the gene level, (ii) equal interchangeability among amino acids, or (iii) replacement drawn from a pool of amino acids at their natural abundances in proteins.

RICHARD HOLMQUIST
Space Sciences Laboratory,
University of California, Berkeley 94720

References and Notes

1. M. O. Dayhoff and L. T. Hunt, *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, D.C., 1972), vol. 5, p. D-355; G. R. Reeck and L. Fisher, *Int. J. Protein Res.* **5**, 109 (1973); T. H. Jukes, R. Holmquist, H. Moise, *Science* **189**, 50 (1975).
2. L. L. Gatlin, *J. Mol. Evol.* **3**, 189 (1974); R. Holmquist and H. Moise, *ibid.* **6**, 1 (1975); A. Cornish-Bowden and A. Marson, *ibid.* **10**, 231 (1977).
3. R. Holmquist, *ibid.* **11**, 349 (1978).
4. A. V. Fowler and I. Zabin, *J. Biol. Chem.* **253**, 5521 (1978).
5. L. McReynolds, B. W. O'Malley, A. D. Nisbet, J. E. Fothergill, D. Givol, S. Fields, M. Robertson, G. G. Brownlee, *Nature (London)* **273**, 723 (1978); A. Ullrich, J. Shine, J. Chirgwin, R. Pictet, E. Tischer, W. Rutter, H. M. Goodman, *Science* **196**, 1313 (1977); J. Shine, P. H. Seeburg, J. Martial, J. D. Baxter, H. M. Goodman, *Nature (London)* **270**, 494 (1977); P. H. Seeburg, J. Shine, J. A. Martial, J. D. Baxter, H. M. Goodman, *ibid.*, p. 486; F. Kafatos, A. Efstratiadis, B. Forget, S. M. Weissman, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5618 (1977).
6. The most recent compilation of the natural abundances of amino acids in proteins is in (*3*) and is, per 61 codons: $Ala_{5.3}$, $Arg_{2.4}$, $Asn_{3.0}$, $Asp_{3.6}$, $Cys_{1.5}$, $Gln_{2.6}$, $Glu_{3.5}$, $Gly_{4.7}$, $His_{1.3}$, $Ile_{3.4}$, $Leu_{4.5}$, $Lys_{4.2}$, $Met_{1.0}$, $Phe_{2.3}$, $Pro_{2.3}$, $Ser_{4.2}$, $Thr_{3.6}$, $Trp_{0.8}$, $Tyr_{2.6}$, $Val_{4.2}$. These averages were obtained from 81 protein families. Other compilations (*1*) of the natural abundances do not differ greatly from the present one and would lead to the same biological conclusions as the text.
7. M. Laird and R. Holmquist, *J. Mol. Evol.* **4**, 261 (1975). See particularly equation 4.
8. E. Margoliash, S. Ferguson-Miller, C. H. Kang, D. L. Brautigan. *Fed. Proc. Am. Soc. Exp. Biol.* **10**, 2124 (1976).
9. This work was supported by NSF grant PCM 76-18627 and NASA grant 05-003-460.

28 September 1978

Binding of *cis*- and *trans*-Dichlorodiammineplatinum(II) to DNA: Evidence for Unwinding and Shortening of the Double Helix

Abstract. *The antitumor drug* cis-*dichlorodiammineplatinum(II) (*cis-*DDP) and the inactive* trans *isomer bind and produce cooperative changes in closed and nicked circular duplex DNA's. Covalent binding of both platinum complexes to the closed circular DNA alters the degree of supercoiling, presumably by disrupting and unwinding the double helix. Electron micrographs show the platinated DNA's to be shortened by up to 50 percent of their original length. At similar ratios of bound platinum per nucleotide, the electrophoretic mobilities of the DNA's in gels containing the dye ethidium bromide are the same for both isomers. The only detectable difference in the binding of the two platinum isomers is an increase in the electrophoretic mobility in nondye gels of closed circular DNA having small amounts of bound* cis-*DDP that is not apparent for the* trans *complex.*

Cis-dichlorodiammineplatinum(II) (*cis*-DDP) is an antitumor drug of clinical importance (*1*, *2*). The *trans* isomer is inactive, however. Since DNA is widely believed to be the primary intracellular binding site of *cis*-DDP, many investigators have looked for differences in the DNA binding properties of the two platinum isomers (*3*, *4*). Previously we showed that *cis*-DDP binds covalently to closed circular PM-2 DNA under conditions (0.2*M* NaCl, 23°C) where *trans*-DDP does not (*5*). Here we report an extension of this work in which the solution conditions more nearly resembled those found intracellularly and in which the actual amount of platinum bound was measured. A remarkable change in the degree of supercoiling of the DNA was found by gel electrophoresis to accompany covalent binding of platinum. In addition, electron microscopy revealed a pronounced and progressive shortening of the DNA with increased platinum binding.

Closed circular DNA's used in this study were isolated from *Escherichia coli* strain K12 W677 cells containing the plasmid pSM1 (*6*). Closed circular, rather than linear, DNA's were employed because of the sensitivity with which small tertiary structural changes can be monitored (*7*) and because their superhelical nature better mimics that of

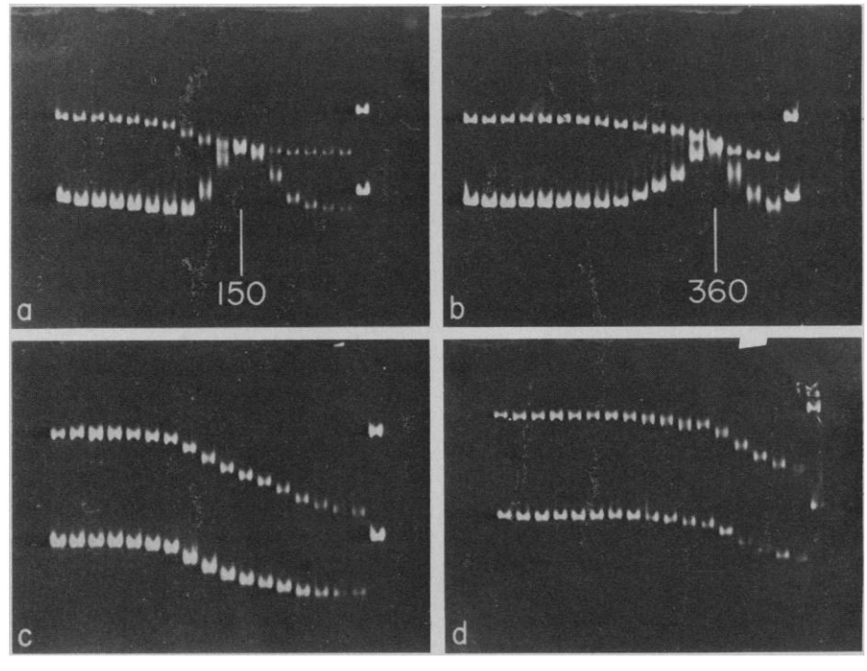

Fig. 1. Electrophoresis in 1 percent agarose gels of nicked and closed circular pSM1 DNA's incubated with (a and c) *cis*-DDP and (b and d) *trans*-DDP as a function of time. Gels contained 0.04*M* tris acetate, 0.02*M* sodium acetate, and 4 m*M* Na-EDTA, *p*H 8.1. The lower (c and d) dye gels also contained ethidium bromide (0.5 μg/ml). The upper (a and b) gels were stained with ethidium bromide (0.5 μg/ml) after electrophoresis. Within each slab, channels correspond to DNA samples (0.2 μg/ml) incubated with platinum reagent for (left to right) 0, 2.5, 5.5, 10, 15, 22, 30, 60, 90, 120, 150, 180, 240, 360, 480, 600, 720, and 0 minutes. Electrophoresis was conducted at 25 V for 15 hours at 4°C.

0036-8075/79/0309-1014$00.50/0