The Influence of Amino Acid Substitutions on the Conformational Energy of Cytochrome *c*†

Paul K. Warme

ABSTRACT: Conformational energies have been evaluated for each of the staggered side-chain conformations associated with the 261 amino acid substitutions known to occur among 60 eucaryotic species. At least 86% of these substitutions can be sterically accommodated (one at a time) within the structure of horse-heart cytochrome *c* resulting from conformational energy refinement. Simultaneous incorporation of all pertinent amino acid substitutions found in eight representative species into the refined horse-heart structure is also shown to be sterically possible, with few exceptions. In two cases (Pekin duck cytochrome with 10 substitutions and *Samia cynthia* cytochrome with 24 substitutions), all substitutions could be readily incorporated, and the total energies associated with their computed structures differed by less than 10 kcal/mol from that of horse-heart cytochrome *c*. In the cytochromes from rattlesnake (22 substitutions), tuna (18 substitutions), and *Neurospora crassa* (36 substitutions), tyrosine could not be substituted for phenylalanine at position 46, within the constraints of the calculations. However, when all of the remaining substitutions were incorporated into these three cytochromes, their computed conformational energies differed by less than 30 kcal/mol from that of horse-heart cytochrome *c*. Between two and four amino acid substitutions cause high energies in the cytochromes from human, baker's yeast, and cotton seed, but all of the remaining substitutions are consistent with a low energy conformation. These results suggest that the structures of homologous proteins may be even more similar than has previously been recognized. Substitutions of all possible amino acid types at the invariant positions (where all eucaryotic cytochromes *c* bear the same amino acid) have revealed some cases where different amino acids can be accommodated, thus demonstrating that the biological constraints on amino acid substitutions are often different from the purely steric constraints investigated in this work.

Amino acid sequences have been determined for the cytochromes *c* isolated from more than 60 different eucaryotic organisms (Margoliash, 1972; Dickerson and Timkovich, 1975). The organisms involved in these sequence studies span the entire plant and animal kingdoms, ranging from yeasts and bacteria to higher plants, insects, fish, reptiles, birds, and mammals. Only 28 of the 104 amino acids of cytochrome *c* are completely invariant among all 60 eucaryotes examined, although fewer than 55% of the amino acids differ when any two species are compared. In spite of these substantial differences in amino acid sequence, the cytochromes *c* from various species appear to be functionally interchangeable in biological assay systems (Byers et al., 1971; Smith et al., 1973), which suggests that their three-dimensional structures must also be very similar. This expectation has been partially borne out by the general similarity among the X-ray structures of horse-heart ferricytochrome *c* at 2.8-Å resolution (Dickerson et al., 1971), tuna-heart ferrocytochrome *c* at 2.45 Å (Takano et al., 1973), *Rhodospirillum rubrum* ferricytochrome c_2 at 2.45 Å (Salemme et al., 1973a,b), and *Micrococcus denitrificans* ferricytochrome c_{550} at 4 Å (Timkovich and Dickerson, 1973). However, because of the substantial effort involved in determining structures by X-ray diffraction methods, it is unlikely that complete X-ray structures will soon be available for the large number of different cytochromes *c* which have been sequenced.

If a refined structure of low energy can be obtained for one of a series of homologous proteins, it should be possible to determine the other homologous structures by means of conformational energy calculations. This approach has been used to predict a structure for α-lactalbumin, based on the assumption that its structure is similar to that of lysozyme (Warme et al., 1974). In the previous paper (Warme and Scheraga, 1975), a low-energy refined structure for horse-heart cytochrome *c* was presented. This paper describes a systematic method for determining which of the 261 amino acid substitutions found among 60 eucaryotic organisms can be readily accommodated (one at a time) within this refined structure. In addition, methods are described for simultaneously inserting all of the amino acid substitutions found in a particular species and for adjusting their side-chain conformations in order to minimize the total energy of the modified protein. The objective of this work is to determine which amino acid substitutions can be accommodated readily within the reference structure of horse-heart cytochrome *c*, and also to discover those which cannot.

Methods

A. General Procedures. The previously refined structure of horse-heart ferricytochrome *c* (Warme and Scheraga, 1975) served as the starting conformation for all of the calculations described in this paper. Amino acid sequence data were taken from the recent compilation of Dickerson and Timkovich (1975), which summarizes all of the known corrections to the originally published sequences. In cases where Asp/Asn or Glu/Gln ambiguities were indicated, the bulkier amide form was assumed. Methylated lysines were treated as regular lysine residues, since the nitrogen atom is likely to be on the external surface of the protein in either

† From the Department of Biochemistry, The Pennsylvania State University, University Park, Pennsylvania 16802. *Received January 28, 1975.* This work was supported by research grants from the Research Corporation, from The Pennsylvania State University, and from the National Science Foundation (BMS74-00090).

case, so that interactions of the ϵ-amino group with other residues should be minimal. Proline substitutions create a special problem in these calculations, since the ϕ angle of proline is normally restricted to values near $-70°$. However, since the backbone conformation was held fixed (in the refined horse-heart conformation) throughout these calculations, proline was considered possible only when the ϕ value at the site of substitution was within 30° of its normal value.

The energy function used for optimization of the side-chain conformational energies included only the nonbonded energy (E_{NB}) and hydrogen bond energy (E_{HB}) contributions (Warme and Scheraga, 1973), since these are the most important forces determining whether a particular conformation is sterically feasible. Furthermore, all interactions of backbone atoms with other backbone atoms were ignored because their contributions to the total energy are constant, due to their fixed positions.

B. Single Amino Acid Substitutions. In this series of energy calculations, all atoms other than the side-chain atoms of the single substituted residue were held fixed at their positions within the refined structure of horse-heart cytochrome *c*. Starting from the all-trans ($\chi_i = -180°$) conformation, the energy associated with each staggered conformation ($\chi_i = -180°, -60°, 60°$) was computed as the sum of the energies of interaction of each side-chain atom of the particular substituted residue with every other atom in the entire protein (including the heme group) lying within the cutoff distances described previously (Warme and Scheraga, 1973). However, to reduce the computation time, the calculation was terminated as soon as any conformation exhibiting a negative total energy was found, since this is a sufficient condition for steric feasibility. For the same reason, the energy calculation was aborted for any given conformation as soon as a total energy in excess of 10 kcal had accumulated, since any such conformation is clearly unfavorable. In the latter case, the search for a favorable side-chain conformation was continued with the next staggered conformation until a conformation exhibiting a negative total energy was found, or until the energies of all 3^N possible staggered conformations had been evaluated (where N is the number of side-chain χ angles). The efficiency of a search procedure such as this may be greatly affected by the order of selection of side-chain conformations; thus, the inner (χ_1) angles were varied most frequently [e.g., if the numbers 1, 2, and 3 are assigned to the conformations $\chi_i = -180°, -60°$, and 60°, respectively, the order of selection for an amino acid having two side-chain dihedral angles would be: ($\chi_1 = 1, \chi_2 = 1$), ($\chi_1 = 2, \chi_2 = 1$), ($\chi_1 = 3, \chi_2 = 1$), ($\chi_1 = 1, \chi_2 = 2$), ($\chi_1 = 2, \chi_2 = 2$), ($\chi_1 = 3, \chi_2 = 2$), etc.], since variations of the outer (χ_2, χ_3, χ_4) angles produce smaller movements of fewer side-chain atoms and thus perturb the overall structure less than variations of the inner angles. Many side chains on the exterior of the protein may take on several favorable conformations, although for our purposes, any conformation exhibiting a negative conformational energy is sufficient to demonstrate steric feasibility, and thus terminates the search procedure.

C. Multiple Amino Acid Substitutions. The simultaneous insertion of all amino acid substitutions found in any given species into the refined structure of horse-heart cytochrome *c* was carried out by a procedure similar to that described for single substitutions, with the exceptions noted here. The substitution closest to the amino terminal was incorporated first, and the energy for each of its staggered side-chain conformations was evaluated in the sequence described in part B. However, the search was not terminated as soon as a negative energy conformation was encountered, but was continued through all 3^N possible staggered conformations, in order to find the one of lowest energy. The efficiency of this search procedure was enhanced by aborting the energy calculation for any given conformation as soon as its accumulated (partial) energy was 10 kcal/mol higher than the total energy of the best conformation previously found (for that substitution). This improves the efficiency of the calculation because the total energy is computed by summing the energies due to a large number of interactions between the side-chain atoms of that residue and all surrounding atoms. These interactions may contribute either positive or negative energies, but if the energy sum at any point exceeds the best previous total energy by more than 10 kcal/mol, it is safe to conclude that the total energy for that conformation will exceed the best previous energy, even without computing the energies of all remaining interactions. After searching all possible conformations, the first substituted amino acid was left in its lowest energy conformation and the remaining substitutions were introduced in similar fashion, proceeding from the amino end to the carboxyl end of the protein.

The lowest energy conformation found for a particular amino acid substitution will depend on the positions of all other atoms in its vicinity, including the side-chain atoms of any other substitutions previously incorporated. Thus, when multiple substitutions are introduced, the order of their insertion may affect their final conformations. For this reason, the process described above was reiterated after all side-chain substitutions had been incorporated, so that each side chain would have the opportunity to optimize its position with respect to all other substituted side chains. In the second iteration of side-chain adjustments, all other substituted residues were held fixed in their best conformations arrived at in the first iteration (or in the second iteration for substitutions at positions of lower sequence number) while the energies of all staggered conformations were reevaluated for each substituted residue in sequence. In every case examined to date, no more than two side-chain conformations have changed during the second iteration, and no changes at all have been observed during a third iteration of side-chain adjustments. This suggests that most of the substitutions are sterically independent of all other substituted amino acids.

After preliminary adjustment of all side-chain conformations by three iterations of the above procedure, a number of minor atomic overlaps still existed in most cases, so a modification of the latter procedure was used to "fine tune" the conformation. Each side-chain angle was incremented by $-10°$, 0°, and 10° from its previous conformation of lowest energy, and after the energies of each of the 3^N possible conformations of this type were evaluated for each substituted residue in turn, the lowest energy conformation was retained. This "fine tuning" procedure was repeated twice, thus allowing deviations of up to 30° (i.e., three increments of 10°) from the staggered conformations, and also providing additional opportunities for each side chain to compensate for the presence of all other substitutions.

Results

A. Single Amino Acid Substitutions. The energies associated with the lowest energy staggered conformations found for each of the 255 non-proline amino acid substitu-

Table I: Energies Associated with Single Amino Acid Substitutions.[a]

	5	10	15	20	25	30	35	40	45	50
Lys = K	1 1 H	H H 1 H		H	H	H		H		1
Arg = R	1	1 2		1				H		
Asp = D	H 1	1		1					1	H
Glu = E	1 H	1	1	H 2	3				1	1
Asn = N	1 1 3	2	2 1	1	1	H 1	1			1
Gln = Q	1 1 1	H H		1 2	1			1 1 H 1		
Ser = S	1 1 1 1	1 1	1	1	1	1 1	1	1 2	1 1 1	
Thr = T	1 2	1 1 1 2	H		H			1 H 3 3	H H 1	
His = H			H		H	H	1			
Tyr = Y						3 1			3 H	
Cys = C		H H	1							
Gly = G H	H 1	1	1 1	H H 1	1	1 H	H H	H	H	
Ala = A	1 1 1 1	1 H	1	1 2 1 3 1		1		H 1	1	1
Val = V	H 3	1 H	H	3	2		1 1	3 2		
Leu = L		1 1 2	3	2		H	H			
Ile = I	1	H 1	1		1		3 1			
Met = M		1								
Pro = P	P				X	H			H	
Phe = F		H					3 3 H		H	
Trp = W							3			

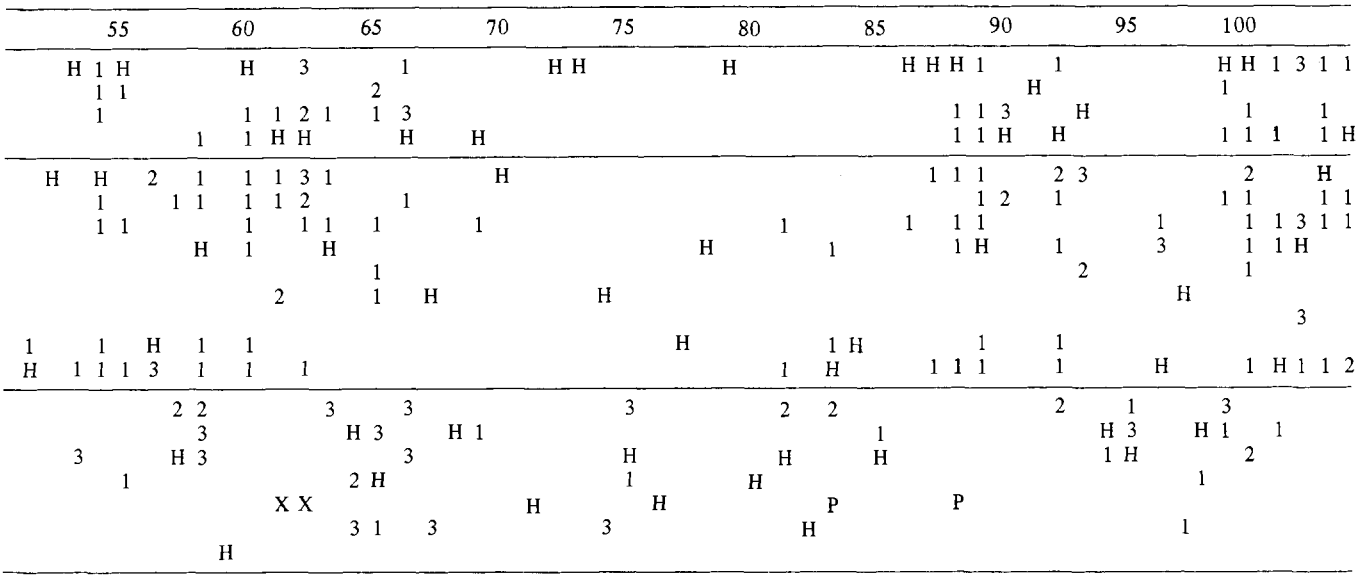
[a] The letters in the outermost columns represent the 20 different types of amino acids, grouped as charged amino acids (above the first horizontal line), neutral hydrophilic amino acids (above the second horizontal line), and hydrophobic amino acids. One vertical column corresponds to each of the 104 sequence positions in mammalian cytochrome *c*, and the H in each column indicates which type of amino acid is found in horse-heart cytochrome *c*. An X denotes a proline residue with an improper ϕ angle, and P denotes a proline with a reasonable ϕ angle ($-70 \pm 30°$). The numbers in each row indicate the energies associated with substitution of that amino acid at a particular sequence

tions that occur in 60 eucaryotic species are summarized in Table I. Although the six proline substitutions that are found in these species were not permitted at this stage of the calculations, those residues having ϕ angles within 30° of the accepted value (−70°) for proline are also indicated. Negative energies were obtained for 188 substitutions (74%), and, thus, these substitutions are sterically favorable. Thirty substitutions (12%) yielded energies between 0 and 10 kcal/mol, which suggests that these are also sterically possible, although further minor adjustments of the side-chain dihedral angles are necessary in these cases. The remaining 37 substitutions (14%) yielded energies greater than 10 kcal/mol. In these cases, further adjustments of the side-chain and/or backbone dihedral angles are definitely required.

It is important to note that only staggered side-chain conformations ($\chi_i = -180°, -60°, 60°$) were allowed at this stage of the calculations, whereas the side-chain angles of natural proteins often deviate by 30° or more from staggered conformations. Preliminary calculations have shown that other substitutions become sterically allowed when deviations of plus or minus 30° from the staggered conformations are permitted. If smaller fluctuations were also permitted, more than 90% of these substitutions would no doubt be sterically allowed. Our results thus represent a lower limit on the percentage of side-chain substitutions which can be accommodated (one at a time) in the refined structure of horse-heart cytochrome *c*. For example, Table I indicates that substitution of phenylalanine for tyrosine at residues 67 and 74 is not allowed, although this would surely be sterically allowed if the actual side-chain angles of the tyrosine residues were used in the calculations, since phenylalanine is smaller than tyrosine, and tyrosine yields a low energy at these positions.

A certain degree of interdependence between one substitution and another might have been expected; for example, when a small amino acid is replaced by a bulkier one, some other amino acid in that region may have to be simultaneously replaced by a smaller one, in order to provide room. Many of the substitutions which proved to be impossible (E = 3 in Table I) may represent cases like this. On the other hand, replacing a bulky amino acid by a smaller one will introduce a cavity in the protein, which might lead to conformational changes, under the influence of many weak attractive forces which conspire to fill in the cavity. The energy calculations presented here are not greatly affected by the existence of cavities, so that substitution of a bulky amino acid by any smaller one will generally be counted as favorable.

All of the amino acid substitutions discussed above are known to occur in one or more of the 60 eucaryotic species included in this study, and thus it may not seem too surprising that they are found to be sterically acceptable. Quite a different result might be expected if we attempted to introduce substitutions that have not yet been observed in any species. Due to the large number of possible substitutions (19 amino acid substitutions at 104 positions would total 1976), an exhaustive study was not possible. However, an indication of the latitude for substitution (based solely on steric criteria) can be gained by attempting all possible amino acid substitutions at the invariant positions, where all eucaryotic cytochromes *c* bear the same identical amino acid. The results of such a study are summarized in Table II. No substitutions could be tolerated at glycine residues 6, 29, 34, 77, and 84, as would be expected on the basis of their tightly packed environments. Several substitutions of small amino acids were allowed at glycine-41, perhaps due to a significant alteration of the conformation in this region during the energy refinement (Warme and Scheraga, 1975). Although some substitutions were found to be possible at proline residues 30, 71, and 76, it appears likely that the special attributes of proline are required at these posi-



position (1 = energy <0; 2 = energy between 0 and 10; 3 = energy >10 kcal/mol). Energies presented in this table are computed for the lowest energy staggered conformation when all other atoms are held fixed at their positions in the refined structure of horse-heart cytochrome *c*. Blanks mean that no substitution of that type occurs among the 60 eucaryotic species examined. The format of this table is patterned after Table X of Dickerson and Timkovich (1975), in order to facilitate comparisons.

tions, for structural reasons. Substitutions at cysteine-17, histidine-18, and methionine-80 were not attempted, because they are covalently linked to the heme group. Tyrosine-48 and tryptophan-59 form hydrogen bonds to one of the propionic acid side chains of the heme group, so substitutions at these positions are unlikely to be tolerated. The roles of the remaining invariant residues are less clearly defined, and in many cases, no definite evidence is available. In each of these cases, the results show that a variety of substitutions can be sterically accommodated within the refined structure of cytochrome *c*, although about half of these can be discounted on the grounds that the substituted amino acid would be in the wrong environment (hydrophilic or hydrophobic). With regard to the remaining (apparently tolerable) substitutions, it seems evident that the biological constraints which prevent these substitutions are quite different from the steric constraints investigated here.

B. Multiple Amino Acid Substitutions. The cytochrome *c* sequences reported for eight species (human, Pekin duck, rattlesnake, tuna, *Samia cynthia* (moth), *Saccharomyces cerevisiae* (baker's yeast), *Neurospora crassa* (mold), and cottonseed) were selected for further study because they represent each of the major classes of organisms (mammals, birds, reptiles, fish, insects, molds, and plants) included among the 60 eucaryotic species examined in the previous section. Although calculations have not been carried out for any other species as of this writing, similar results are anticipated.

The total energies obtained at various stages of side-chain adjustment are shown in Table III for each of these eight species. Although only the nonbonded and hydrogen bonded energy contributions were taken into account during the process of optimizing the side-chain conformations, the energies given in Table III include the electrostatic and rotational energies as well, in order to facilitate comparison with the total energy of refined horse-heart cytochrome *c* (−504 kcal/mol) reported previously (Warme and Scheraga, 1975). After "coarse" adjustment (allowing only staggered side-chain conformations), only Pekin duck cytochrome *c* had an energy similar to that of horse-heart, while the other seven had energies ranging from 84 to 181 kcal/mol higher. After "fine" adjustment (allowing ±10, 20, or 30° deviations from staggered conformations), *Samia cynthia* cytochrome *c* had an energy only 9 kcal/mol higher than that of horse-heart, and the total energy of Pekin duck cytochrome *c* was slightly lower than the reference energy of horse-heart cytochrome *c*. The other six cytochromes all exhibited energies significantly lower than those obtained after "coarse" adjustment, but their energies still ranged from 28 to 96 kcal/mol higher than that of horse-heart cytochrome *c*, due to the presence of one or more atomic overlaps in each case.

The differences between the energies calculated for each amino acid residue of horse-heart cytochrome *c* and the corresponding energies for these eight cytochrome *c* variants after "fine" adjustment are plotted in Figure 1. Almost all of the energy differences are less than 10 kcal/mol. In many cases, an increase in the energy of one amino acid is compensated by a decrease in the energy of an adjacent amino acid. The latter observation may simply reflect changes in the interactions of the neighboring residues, but it might also be taken as an indication that a stable conformation depends on the total energy of the entire protein more than it depends on the energies of particular amino acids. In those few instances where the energy difference exceeds 10 kcal/mol, it appears likely that further adjustments of side-chain and/or backbone dihedral angles will be necessary to achieve a stable overall conformation. However, in some of these cases, the energy for a particular amino acid found in horse-heart cytochrome was very negative, while the amino acid substituted at that position also gave a negative, but nonetheless higher, energy; therefore,

Table II: Energies Associated with Amino Acid Substitutions at Invariant Positions.[a]

	K	R	D	E	N	Q	S	T	H	Y	C	G	A	V	L	I	M	P	F	W	Environment
Gly- 6											–	–						–			Internal
Phe-10	–2		–1		–2		–1	–1			–	–	1		2		1	–			Internal
Gln-16	1	1	1	1	1	1	1	1	1	–1	–	–	1	–1	–1	–1	–1	–	–1	–2	External
Cys-17	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Heme attachment
His- 18	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Heme ligand
Gly-29											–	–						–			Surface
Pro-30							–1				–	–	1					–			Internal
Leu-32			–1		–1		–1	–1		1	–	–	1	1		2	1	–			Internal
Gly-34											–	–						–			Surface
Arg-38	2	2	2	2	2	2	1	1			–	–	2					–			Surface
Gly-41			2				1				–	–	1					–			Surface
Tyr-48	–2	–1	–1	–1	–1	–2	–1	–1			–	–	1	1		1	1	–			Internal
Asn-52	–1		–1	–1	–1	–1	–1	–1	–2		–	–	1	1		1	2	–			Internal
Trp-59	–2	–2		–1		–1	–1				–	–	1				2	–			Internal
Leu-68	–2		–1	–1	–1	–1	–1	–1	–1	1	–	–	2	2	2		1	–	1		Internal
Asn-70	1	1	1	1	2	1	1	2			–	–	1	–1	–1	–1	–1	–			External
Pro-71			–1		–1		–1	–1	–1		–	–	1					–			Internal
Lys-73	1	1	1	1	1	1	1	2	1	–1	–	–	1		–1		–1	–	–1	–1	External
Pro-76	1	1	1	1	1	1	1	1	1	–1	–	–	1	–1	–1	–1	–1	–	–1	–1	External
Gly-77											–	–									External
Thr-78			–1	–1	–1	–1	–1				–	–	1	1		1		–			Internal
Lys-79	1	1	1	2	2	1	1	1	2	–2	–	–	1	–1	–1	–2	–1	–	–2	–2	External
Met-80	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Heme ligand
Phe-82	–1	–1	–1	–1	–1	–1	–1	–1			–	–	1	1	1	2	1	–			Internal
Gly-84											–	–						–			Surface
Arg-91			1	1			1				–	–	1					–			Surface

[a] The energies given in this table are computed for the lowest energy staggered conformations, when all other atoms are held fixed at their positions in the refined structure of horse-heart cytochrome *c*. The energies are represented by numbers according to the following code: ±1 means a negative (favorable) energy, ±2 means an energy between 0 and 10 kcal/mol, and blank means that no conformation with energy less than 10 kcal/mol was found. Calculations were omitted for the trivial case of glycine substitution, and substitutions of proline and cysteine were not allowed because these amino acids normally play a special role in defining protein structures. Substitutions for cysteine-17, histidine-18, and methionine-80 were also omitted, since these residues play critical roles in cytochrome *c*. All of these disallowed cases are designated by minus signs in the table. Negative numbers elsewhere in the table denote substitutions which would not be expected to occur because this would place a hydrophobic amino acid in an aqueous environment or would place a hydrophilic amino acid in the nonpolar interior of the protein. The one-letter amino acid code is defined in Table I.

Table III: Energies after Side-Chain Adjustments on Cytochromes *c* from Eight Eucaryotic Species.[a]

Species	No. of Substitutions	Total Energy (kcal/mol) A. After Coarse Adjustment	B. After Fine Adjustment	C. Omitting Specified Substitutions	Substitutions Omitted
Human	12	–418	–441	–500	Y46,D62
Pekin duck	10	–501	–505		
Rattlesnake	22	–384	–428	–484	Y46
Tuna	18	–420	–462	–517	Y46
Samia cynthia	24	–412	–495		
Saccharomyces	39	–352[b]	–409[b]	–465[b]	Y46,I53
Neurospora	36	–350	–476	–531	Y46
Cottonseed	40	–323[c]	–408[c]	–530[c]	A24,T42,T43,Y46

[a] The total energy computed for refined horse-heart cytochrome *c* is –504 kcal/mol (Warme and Scheraga, 1975). [b] I53 energy omitted. [c] A24 energy omitted.

the energy difference was large, even though the substituted residue was sterically favorable. One example of this occurs when glutamic acid residue 4 (energy = –11.5 kcal/mol) in horse-heart is replaced by lysine in *Saccharomyces, Neurospora* and cottonseed cytochromes, yielding a positive energy difference of 10.5–13.5 kcal/mol. In spite of this apparently unfavorable energy difference, lysine must be considered sterically feasible at position 4, even though it is significantly less favorable than glutamic acid.

Substitution of tyrosine for phenylalanine at position 46 contributes a large positive energy (~54 kcal/mol divided among residues 29, 30, and 46) in each of the six species where it is found (human, rattlesnake, tuna, *Saccharomyces, Neurospora,* and cottonseed). This high energy is entirely due to overlaps of the hydroxyl oxygen atom of tyrosine-46 with the C^{α} and C′ atoms of glycine-29, in addition to close contacts of this same oxygen atom with the N and C^{δ} atoms of proline-30. In each of the six species where tyrosine is found at this position, the same conformation of lowest energy was calculated but apparently we have not allowed sufficient conformational freedom to permit any other conformation of lower energy. As discussed previously (Warme and Scheraga, 1975), the position of phenylalanine-46 had to be altered markedly during the energy refinement of horse-heart cytochrome *c* in order to reach a low-energy conformation in this local region. Thus, there is reason to suspect that its position is not optimal, possibly due to errors in the X-ray coordinates somewhere in this region. When phenylalanine was reinstated at residue 46, the total energies of rattlesnake, tuna, and *Neurospora* cytochromes *c* were all within 20 kcal of the energy calculated for horse-heart cytochrome *c* (Table III).

Several other substitutions also led to atomic overlaps that contributed positive energies in excess of 10 kcal/mol. Substitution of isoleucine for lysine at position 53 in *Saccharomyces* and substitution of alanine for glycine at posi-

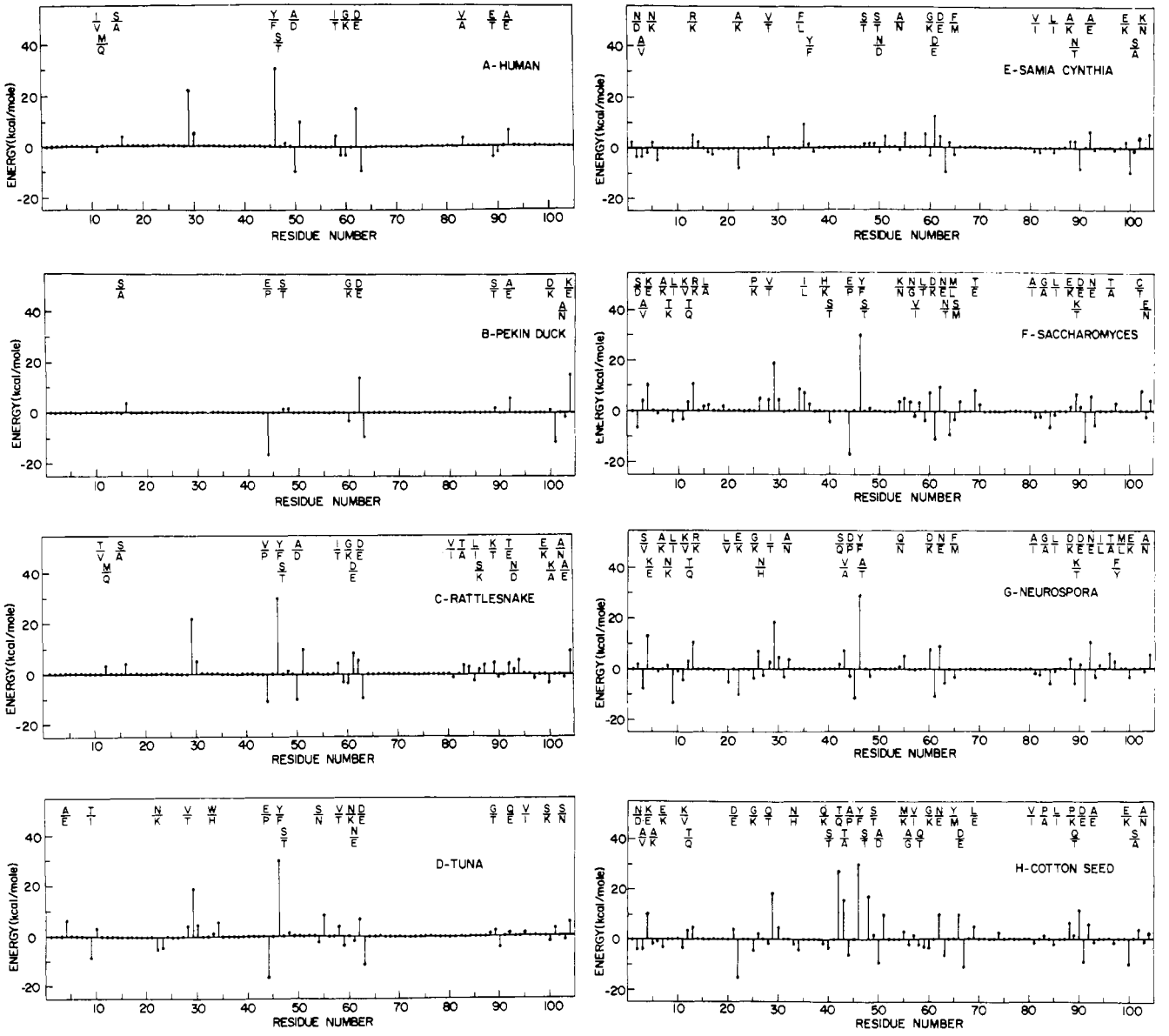


FIGURE 1: For each amino acid in the computed structure of the specified variety of cytochrome *c*, the difference between its energy and the energy of the amino acid at the corresponding position in horse-heart cytochrome *c* is shown. Near the top of each graph, the positions of amino acid substitutions are indicated, using the one-letter code defined in Table I. The upper letter of each pair denotes the amino acid found in the given species, while the lower letter denotes the corresponding amino acid in horse-heart cytochrome *c*. Note that energy differences do not necessarily coincide with positions where substitutions occur.

tion 24 in cottonseed invariably produced energies in excess of 100 kcal/mol and, thus, these substitutions were omitted in the calculations. Backbone conformational changes will probably be required before these two substitutions can be accommodated. Aspartic acid is quite unfavorable at position 62 in human cytochrome *c*, as is threonine at positions 42 and 43 in cottonseed cytochrome *c*. When these substitutions are omitted from human, *Saccharomyces*, and cottonseed cytochrome *c*, all of them exhibit energies within 40 kcal/mol of the energy computed for horse-heart cytochrome *c* (Table III).

Table IV lists the most favorable side-chain conformations computed for each substitution incorporated into the eight species varieties of cytochrome *c* studied in this work. Taken within the context of the assumptions and constraints affecting these calculations, these lists constitute predictions of the structures of the cytochromes *c* from these species.

Discussion

Several limitations of these calculations must be mentioned, in order to place these results in proper perspective. First of all, the refined structure of horse-heart cytochrome *c* used in this work contains some regions of uncertainty. The accuracy of this structure depends, in turn, upon the accuracy of the X-ray coordinates, which must be considered provisional due to the fact that they are based on a 2.8-Å resolution electron density map (R. E. Dickerson, personal communication). It is hoped that the accuracy of the computed structure for horse-heart cytochrome *c* can be improved by taking advantage of further information provided by the 2-Å resolution structure of tuna cytochrome *c*, which should be available soon. Another limitation of the calculations is that the effects of amino acid insertions at the amino terminal (in *Samia cynthia*, *Saccharomyces*, *Neurospora*, and cottonseed cytochromes *c*) were not taken into account, since no X-ray coordinates are available for

Table IV: Side-Chain Conformations for Substituted Residues in Eight Representative Eucaryotic Cytochromes *c*.[a]

HUMAN					
11	I	-80	-40		
12	M	-160	-150	150	
15	S	-180	-50		
46	Y	-180	-80	90	
47	S	90	-170		
50	A				
58	I	-60	-70		
60	G				
62	D	-90	-170		
83	V	-180			
89	E	-180	70	-90	
92	A				
PEKIN DUCK					
15	S	-180	-50		
44	E	-70	-40	-70	
47	S	90	-170		
60	G				
62	D	-80	-180		
89	S	30	150		
92	A				
100	D	-160	80		
103	A				
104	K	-30	160	-80	60
SACCHAROMYCES					
2	S	-90	-30		
3	A				
4	K	-180	170	30	50
7	A				
8	T	-30	150		
9	L	-70	170		
11	K	160	-150	80	-80
12	T	-70	-30		
13	R	-80	-180	-30	60
15	L	170	60		
25	P				
28	V	160			
35	I	170	80		
39	H	-90	-70		
40	S	-180	-160		
44	E	-70	-40	-70	
46	Y	-180	-80	90	
47	S	90	-170		
54	K	-170	-170	30	-70
56	N	-90	-70		
57	V	90			
58	L	-90	-160		
60	D	-50	50		
62	N	-150	-150		
63	N	-50	-30		
64	M	-160	40	90	
65	S	30	150		
69	T	-50	-40		
81	A				
83	G				
85	L	-90	160		
88	E	-160	30	70	
89	K	-160	-150	40	70
90	D	160	80		
92	N	170	50		
96	T	-30	150		
102	S	-180	-80		
103	E	-150	30	70	
COTTON SEED					
2	N	-50	70		
3	A				
4	K	-180	170	30	50
5	A				
7	E	-170	150	170	
11	K	160	-150	80	-80
12	T	-70	-30		
21	D	-180	60		
25	G				
28	Q	170	80	90	
33	N	-160	80		
39	Q	-70	-50	-70	
40	S	-180	-160		
42	T	-60	-80		
43	T	160	70		
44	A				
46	Y	-180	-80	90	
47	S	90	-170		
49	S	40	-150		
50	A				
55	N	50	-160	50	
56	A				
57	V	80			
58	Q	-60	-60	-70	
60	G				
62	N	-150	-150		
65	Y	50	70	-150	
66	D	-150	-60		
69	L	-180	90		
81	V	-50			
83	P				
85	L	-90	150		
88	P				
89	Q	-180	70	-90	
90	D	-80	-170		
92	A				
100	E	-70	170	90	
101	S	60	40		
103	A				
TUNA					
4	A				
9	T	-30	150		
22	M	-70	70		
28	V	160			
33	M	70	-80		
44	E	-70	-40	-70	
46	Y	-180	-80	90	
47	S	90	-170		
54	S	60	-90		
58	V	90			
60	N	-60	40		
61	N	-60	70		
62	D	-150	-150		
89	G				
92	Q	-180	-180	-60	
95	V	170			
100	S	-50	90		
103	S	60	40		
RATTLESNAKE					
11	T	-30	-90		
12	M	-160	-150	150	
15	S	-180	-50		
44	V	-180			
46	Y	-180	-80	90	
47	S	90	-170		
50	A				
58	I	-60	-70		
60	G				
61	D	-60	80		
62	D	-90	-170		
81	V	-50			
83	T	-30	-90		
85	L	-70	160		
86	S	-170	-30		
89	K	-180	-170	40	70
92	T	-40	150		
93	N	-50	-30		
100	E	-70	-30	-70	
101	K	-180	-180	-50	-170
103	A				
104	A				
SAMIA CYNTHIA					
2	N	-50	70		
3	A				
5	N	-150	-70		
13	R	-180	60	-180	-60
22	A				
28	V	160			
35	P	-170	160		
36	Y	170	70	60	
47	S	90	-170		
49	S	40	-150		
50	N	50	-60		
54	A				
60	G				
61	D	-30	-70		
62	D	-80	-180		
65	F	50	70		
81	V	-50			
85	L	-70	160		
88	A				
89	N	170	60		
92	A				
100	E	-70	170	90	
101	S	60	40		
103	K	-150	90	70	-150
NEUROSPORA					
3	S	-30	-80		
4	K	-180	170	30	50
7	A				
8	N	-180	30		
9	L	-70	170		
11	K	160	-160	80	-70
12	T	-70	-30		
13	R	-80	-180	-30	60
20	L	-160	150		
22	E	-180	170	-30	
25	G				
26	N	-90	-50		
28	I	-70	150		
31	A				
42	S	-160	-150		
43	V	50			
44	D	60	-60		
46	Y	-180	-80	90	
47	A				
54	Q	-80	-30	-60	
60	D	-50	50		
62	N	-150	-160		
65	F	50	70		
81	A				
83	G				
85	L	-90	170		
88	D	-170	70		
89	K	-160	-150	40	70
90	D	160	80		
92	N	-60	-30		
94	I	-70	-50		
96	T	-30	-90		
97	F	-170	60		
98	M	-70	-170	90	
100	E	-70	-30	-70	
103	A				

[a] From left to right, the columns indicate the sequence number, the amino acid substituted at that position (see Table I for definitions of one-letter code) and the computed side-chain dihedral angles (χ_1,χ_2,χ_3,χ_4).

these amino acids. The lowest energy side-chain conformations computed by the incremental energy minimization method used in this study do not represent the precise minimum of energy because the smallest variation allowed for any angle was 10°; thus, the true minimum is probably within 5° of the computed value for most angles. Deviations of more than 30° (i.e., three increments of 10° each) from the staggered conformations were not allowed, although the minimum-energy conformation might lie outside this range in some cases. It is also quite possible that a lower energy conformation could be obtained if the backbone dihedral angles and the side-chain dihedral angles of the unsubstituted side chains were allowed to vary, as well as the side chains of the substituted amino acids. Further calculations permitting such variations may provide information about the magnitudes of the conformational changes which are required in order to accommodate the relatively few amino acid substitutions which were impossible in the absence of these extra degrees of freedom. The energies computed in this work included only the nonbonded energy (E_{NB}) and hydrogen bond energy (E_{HB}), but did not take into account the electrostatic energy (E_{EL}) or rotational energy (E_{ROT}). Because of the fact that only staggered conformations (or increments of ±10, 20, or 30° from the staggered conformation) were allowed, it is to be expected that the rotational energy is near the minimum possible value. However, neglect of electrostatic energies might affect our results, and in borderline cases, some conformation other than the one which resulted here might be more favorable. Thus, if the acceptability of any particular substitution hinges on energy differences of magnitude less than a few kilocalories, our results might be misleading. In spite of these limitations, the systematic approach used in this work has the advantage of simplicity, thus allowing a more accurate description of the methods and results, as well as their important implications.

The steric feasibility of any particular substitution will generally be affected by any other substitutions which occur in a given species. A priori, one might expect that a particular amino acid substitution found in rattlesnake would not necessarily fit within the horse structure, since other amino acid substitutions which occur in the rattlesnake structure might be a necessary prerequisite. The results of our studies on single amino acid substitutions seem to indicate that this situation arises rather infrequently in cytochrome *c*. Statistical studies of evolutionary variations in cytochrome *c* amino acid sequences have suggested that as few as four to ten amino acids are capable of substitution within any given species at a given time on the evolutionary scale (Fitch and Markowitz, 1970; Fitch, 1971). The set of concomitantly variable codons (covarions) generally changes with each amino acid substitution introduced, thus allowing different amino acids to join the pool of covarions. Over a long period of evolutionary time, the set of covarions encompasses a large fraction of the entire amino acid sequence, thereby explaining the observation that about 75% of the amino acids in cytochrome *c* are capable of variation, when all known sequences are compared. If this covarion hypothesis is accepted, then our observation that more than 85% of the known amino acid substitutions found among eucaryotic species can be sterically accommodated (one at a time) within the structure of horse-heart cytochrome *c* must lead to the conclusion that the biological constraints are often different from the purely steric constraints which are investigated in this work. This point is further reinforced by the results of the calculations dealing with amino acid substitu-

tions at invariant positions. In many cases, it was observed that a variety of substitutions could be sterically accommodated, although they apparently are not biologically acceptable. A good example is the case of glutamine 16, which projects into the solution near the heme group. This position could be occupied by any of the amino acids tested, although hydrophobic side chains would not be favored by the aqueous environment. The fact that glutamine is invariant at this position suggests that it must be involved in interactions with other proteins in the electron transport chain. The same arguments pertain to asparagine-70, lysine-73, lysine-79, and arginine-91. A more likely interpretation of the constancy of phenylalanine-10, leucine-32, asparagine-52, leucine-68, threonine-78, and phenylalanine-82 is that these amino acids play a role in stabilizing a compact, space filling tertiary structure for cytochrome *c* (by participating in internal hydrogen bonds or van der Waals interactions), since these residues are all in the interior of the protein.

Several lines of argument have been advanced to support the hypothesis that the tertiary structures of all cytochromes *c* are very similar to one another. A strong case can be made on the basis of sequence homology. The sequences for some species (e.g., man and monkey) differ at only one or two positions, so it is not difficult to believe that they have similar structures. However, there seems to be an almost continuous gradation in the number of differences up to the level of more than 60 differences (when additions and deletions at the amino and carboxyl ends of the chain are accounted for). When there are this many differences in amino acid sequence, the assumption of structural similarity becomes much less plausible. Only 28 amino acids are completely invariant among 60 eucaryotic species (Dickerson and Timkovich, 1975) and it seems likely that these amino acids are important because of their involvement in the electron transfer function of cytochrome *c* or for preserving favorable interactions with cytochrome reductase and oxidase or else for producing and maintaining the appropriate tertiary structure of the protein. Many other amino acid substitutions which involve replacement of one polar or nonpolar amino acid for another with similar properties have been called "conservative" substitutions, on the premise that such a slight change should not impair the function of that particular amino acid (Margoliash and Schejter, 1966). The difficulty with this classification is that the precise function of each amino acid in cytochrome *c* is not yet clearly understood, so we should reserve judgement on this question of whether the function of one amino acid in a particular environment can be served by any other type of amino acid. Normally, one would consider substitution of phenylalanine for tyrosine to be a "conservative" substitution, although a major reappraisal of the mechanism of electron transfer was required when this substitution was recently detected at positions 67 and 74 (Takano et al., 1973). Moreover, the conformations of residues which are not critical for the function of cytochrome *c* might differ significantly from one species to the next, provided that the resulting structural perturbations are not so large that they disturb the positions of the mechanistically important residues excessively. For these reasons, the arguments for structural similarity based on sequence homology cannot be considered conclusive and leave us with questions concerning the degree of similarity.

A second strong argument in favor of structural similarity for all eucaryotic cytochromes *c* is based on their functional equivalence in in vitro electron transport systems. The cytochromes from many species, including insects, fish, birds, and mammals, are able to transfer electrons to bovine cytochrome oxidase at similar rates (Smith et al., 1973), in spite of many differences in their amino acid sequences. Since the exact mechanism of electron transfer and the precise mode of interaction between cytochrome *c* and cytochrome oxidase are not known, it is difficult to say how similar the structures of these various cytochromes *c* must be in order to perform this function. It seems likely that only a fraction of the surface area of cytochrome *c* interacts directly with cytochrome oxidase, so the remainder of the protein could presumably differ quite significantly from one species to the next, without necessarily preventing normal electron transfer to cytochrome oxidase.

X-Ray diffraction studies have provided the most compelling evidence for structural similarities among diverse classes of cytochrome *c*, although the data are still quite fragmentary, and higher resolution structures will be required before many doubts are eliminated. When bonito cytochrome *c* was compared with the horse-heart structure by difference map techniques, no major differences in the overall pattern of chain folding were detected (Dickerson et al., 1971). Close similarities between the cytochrome c_2 from *Rhodospirillum rubrum* and horse-heart cytochrome *c* have been described by Salemme et al. (1973b), but many major differences are also evident. The 4-Å resolution structure of *Micrococcus denitrificans* cytochrome *c* appears to resemble both horse-heart cytochrome *c* and *R. rubrum* c_2 in different regions, but many features of the *M. denitrificans* structure are still difficult to discern at this resolution (Timkovich and Dickerson, 1973). The 2-Å resolution structure of tuna cytochrome *c* will be known in the near future (according to Dickerson and Timkovich, 1975), and a detailed comparison with horse-heart cytochrome *c* should permit a much more definitive statement concerning the extent of structural similarity among the eucaryotic cytochromes *c*. However, in view of the large number of amino acid substitutions found in certain species, it is likely that many minor (and possibly some major) differences in conformation exist in some of the more distantly related species.

All of the preceding lines of evidence fail to provide a clear answer to our questions concerning the precise degree of structural similarities among the eucaryotic cytochromes *c*. Are only the side-chain conformations different near the sites of substitution or are backbone conformational changes also required, at least in a few cases? Must the conformations of unsubstituted side chains be altered significantly in order to compensate for the presence (or absence) of certain atoms at the sites of amino acid substitution? Lewis and Scheraga (1971) have shown that the helix probability profiles for 27 different eucaryotic species are consistent with preservation of three helical regions in all species except three (fruit fly, screw-worm fly, and yeast), in which only two helical regions are predicted. The results presented in this paper suggest that the structures of the various eucaryotic cytochromes *c* may be even more similar than hitherto suspected. Even when all of the backbone atoms and all of the unsubstituted side-chain atoms are maintained fixed in the same positions as in the refined structure of horse-heart cytochrome *c*, as many as 35 out of 36 amino acid substitutions in *Neurospora* cytochrome *c* can be readily accommodated; moreover, the total optimized energy is even lower than that calculated for horse-heart cytochrome *c*.

The side-chain conformations listed in Table III (in context with the basic assumption that the remainder of the protein is identical with horse-heart cytochrome *c*) may be considered to be predicitons of the tertiary structures of these eight cytochromes *c*. It will be interesting to compare the calculated side-chain conformations for tuna cytochrome *c* with the actual conformations of these residues when the X-ray structure becomes available. The results obtained in this work indicate that 17 of the 18 substitutions found in tuna cytochrome *c* can be incorporated readily into the refined structure of horse-heart cytochrome *c*, and this provides encouragement for future attempts to apply the information gained from the more accurately determined X-ray structure of tuna cytochrome *c* to the task of further refining the structure of horse-heart cytochrome *c* by conformational energy calculations.

Acknowledgments

Thanks are due to Dr. H. A. Scheraga for his encouragement during this work, to Dr. R. E. Dickerson for supplying a copy of his review article prior to publication, and to Dr. E. Margoliash for helpful discussions of this work.

References

Byers, V., Lambeth, D., Lardy, H. A., and Margoliash, E. (1971), *Fed. Proc., Fed. Am. Soc. Exp. Biol. 30,* 1286.

Dickerson, R. E., Takano, T., Eisenberg, D., Kallai, O. B., Samson, L., Cooper, A., and Margoliash, E. (1971), *J. Biol. Chem. 246,* 1511.

Dickerson, R. E., and Timkovich, R. (1975), *Enzymes, 3rd Ed.* (in press).

Fitch, W. M. (1971), *J. Mol. Evol. 1,* 84.

Fitch, W. M., and Markowitz, E. (1970), *Biochem. Genet. 4,* 579.

Lewis, P. N., and Scheraga, H. A. (1971), *Arch. Biochem. Biophys. 144,* 576.

Margoliash, E. (1972), *Harvey Lect. 66,* 177.

Margoliash, E., and Schejter, A. (1966), *Adv. Protein Chem. 21,* 113.

Salemme, F. R., Freer, S. T., Xuong, N. H., Alden, R. A., and Kraut, J. (1973a), *J. Biol. Chem. 248,* 3910.

Salemme, F. R., Kraut, J., and Kamen, M. D. (1973b), *J. Biol. Chem. 248,* 7701.

Smith, L., Nava, M. E., and Margoliash, E. (1973), *Oxidases Relat. Redox Syst., Proc. Int. Symp., 2nd,* 629.

Takano, T., Kallai, O. B., Swanson, R., and Dickerson, R. E. (1973), *J. Biol. Chem. 248,* 5234.

Timkovich, R., and Dickerson, R. E. (1973), *J. Mol. Biol. 79,* 39.

Warme, P. K., Momany, F. A., Rumball, S. S., Tuttle, R. W., and Scheraga, H. A. (1974), *Biochemistry 13,* 768.

Warme, P. K., and Scheraga, H. A. (1973), *J. Comput. Phys. 12,* 49.

Warme, P. K., and Scheraga, H. A. (1974), *Biochemistry 13,* 757.

Warme, P. K., and Scheraga, H. A. (1975), *Biochemistry,* preceding paper in this issue.

Cobalt–Cytochrome *c*. I. Preparation, Properties, and Enzymic Activity†

L. Charles Dickinson and James C. W. Chien*

ABSTRACT: An improved procedure for the preparation of cobalt-cytochrome *c* has been developed. Various factors influencing the cobalt insertion process are discussed. The optical spectra of cobalt-cytochrome *c* suggest a six-coordinated species. The spectral shifts occurring with oxidation-reduction are compared with those observed for deoxycobaltohemoglobin and ferrocytochrome *c* and attributed to the effect of d_{z^2} electron on stereoelectronic interactions between the axial ligands and the porphyrin π systems. Cobalt-cytochrome *c* has $E_{m,7} = -140 \pm 20$ mV as compared to an $E_{m,7}$ of +250 mV for ferrocytochrome *c*. An explanation for this negative $E_{m,7}$ is offered. Cobaltocytochrome *c* is oxidized by cytochrome oxidase at about 45% of the rate for native cytochrome *c*. On the other hand cobalticytochrome *c* was not reduced by microsomal NADH or NADPH cytochrome *c* reductase nor by mitochondrial NADH or succinate cytochrome *c* reductase. It appears that the integrity of the reductase binding site is destroyed and the oxidase binding site has been modified by cobalt substitution.

The functional properties of a metalloenzyme are a direct manifestation of its primary sequence, protein conformation, the metal ion, and the prosthetic group. Even though X-ray diffraction can provide a detailed image of an enzyme molecule, it is less useful in answering questions concerning metal ion specificities. For instance, why does nature select Fe for hemoglobin but Co for vitamin B_{12} or Zn to activate carboxypeptidase but Mg to activate enolase? Recently, there is a growing interest in the study of metal substituted enzymes to understand structure-function relationships. A case in point is that of cobalthemoglobin, an allosteric analog of hemoglobin. Comparisons of the properties of these two molecules (Hoffman and Petering, 1970;

† From the Department of Chemistry, University of Massachusetts, Amherst, Massachusetts 01002. *Received March 3, 1975.* This work was supported in part by Grant No. HL-14270 from the U.S. Public Health Service.