ACKNOWLEDGMENTS

We appreciate the careful review of the manuscript by the referees.

REFERENCES

Amouyal, M., & Buc, H. (1987) *J. Mol. Biol. 195*, 795–808.
Auble, D. T., & deHaseth, P. L. (1988) *J. Mol. Biol. 202*, 471–482.
Auble, D. T., Allen, T. L., & deHaseth, P. L. (1986) *J. Biol. Chem. 261*, 11202–11206.
Ayers, D. G., Auble, D. T., & deHaseth, P. L. (1989) *J. Mol. Biol.* (in press).
Breslauer, K., Remeta, D. P., Chou, W.-Y., Ferrante, R., Curry, J., Zaunczkowski, D., Snyder, J. G., & Marky, L. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8922–8926.
Bruzik, J. P., Auble, D. T., & deHaseth, P. L. (1987) *Biochemistry 26*, 950–956.
Burkhoff, A., & Tullius, T. (1987) *Cell 48*, 935–943.
Chamberlin, M. J., Nierman, W. C., Wiggs, J., & Neff, N. (1979) *J. Biol. Chem. 254*, 10061–10069.
Coll, M., Frederick, C. A., Wang, A. H.-J., & Rich, A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8385–8389.
Hertzberg, R. P., & Dervan, P. B. (1982) *J. Am. Chem. Soc. 104*, 313–315.
Hertzberg, R. P., & Dervan, P. B. (1984) *Biochemistry 23*, 3934–3945.
Klevit, R. E., Wemmer, D. E., & Reid, B. R. (1986) *Biochemistry 25*, 3296–3303.
Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., & Dickerson, R. E. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1376–1380.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Marky, L. A., Blumenfeld, K. S., & Breslauer, K. J. (1983) *Nucleic Acids Res. 11*, 2857–2870.
McClure, W. R. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5634–5638.
McClure, W. R. (1985) *Annu. Rev. Biochem. 54*, 171–204.
Nelson, H. C. M., Finch, J. T., Luisi, B. F., & Klug, A. (1987) *Nature 330*, 221–226.
Schultz, P. G., & Dervan, P. B. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 6834–6837.
Schultz, P. G., & Dervan, P. B. (1984) *Biomol. Struct. Dyn. 1*, 1133–1147.
Siebenlist, U., Simpson, R. B., & Gilbert, W. (1980) *Cell 20*, 269–281.
Snounou, G., & Malcolm, A. D. B. (1983) *J. Mol. Biol. 167*, 211–216.
Van Dyke, M. W., & Dervan, P. B. (1983) *Nucleic Acids Res. 11*, 5555–5567.
Van Dyke, M. W., Hertzberg, R. P., & Dervan, P. B. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 5470–5474.
von Hippel, P. H., Bear, D. G., Morgan, W. D., & McSwiggen, J. A. (1984) *Annu. Rev. Biochem. 53*, 389–446.
Wilson, W. D., Wang, Y., Krishnamoorthy, C. R., & Smith, J. C. (1985) *Biochemistry 24*, 3991–3999.
Yoon, C., Prive, G. G., Goodsell, D. S., & Dickerson, R. E. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6332–6336.
Youngquist, R. S. (1988) Ph.D. Thesis, California Institute of Technology.
Youngquist, R. S., & Dervan, P. B. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2565–2569.
Zimmer, C., & Wahnert, U. (1986) *Prog. Biophys. Mol. Biol. 47*, 31–112.
Zunino, F., & DiMarco, A. (1972) *Biochem. Pharmacol. 21*, 867–873.

Comparison of the Data, Analysis, and Results of X-ray Absorption Studies of Cytochrome *c* Oxidase

L. Powers* and Brian M. Kincaid‡
AT&T Bell Laboratories, Murray Hill, New Jersey 07974
Received February 4, 1988; Revised Manuscript Received November 14, 1988

ABSTRACT: Differences in the methods of analysis of X-ray absorption data used by Powers et al. [Powers, L., Blumberg, W. E., Chance, B., Barlow, C., Leigh, J., Jr., Smith, J., Yonetani, T., Vik, S., & Peisach, J. (1979) *Biochim. Biophys. Acta 547*, 520–538; Powers, L., Chance, B., Ching, Y., & Angiolillo, P. (1981) *Biophys. J. 34*, 465–498] and Scott et al. [Scott, R., Schwartz, J., & Cramer, S. (1986) *Biochemistry 25*, 5546–5555] are clarified. In addition, we compare the X-ray absorption data and results for resting cytochrome *c* oxidase reported by both groups using the same analysis method and conclude apart from any assumptions that the data are not identical.

In order to clarify the differences and disagreements between our results (Powers et al., 1979, 1981) and those of Scott et al. (1986) (SSC) apart from any assumptions in the analysis methods, this paper compares directly the data and results of X-ray absorption studies of resting cytochrome *c* oxidase from both groups. We point out that SSC have not represented correctly the use of the site modeling method (Powers et al., 1981). Taking into account the reduction in number of degrees of freedom in Fourier-filtered EXAFS data, application of fitting methods to the published data of both groups shows that the data are not identical. Comparison of the analysis methods of both in view of the assumptions inherent in each clarifies the differences in the proposed respective models.

Cytochrome *c* oxidase contains four metal redox sites, two copper and two iron. One iron and copper pair functions as an electron reservoir, Cu_a and Fe_a, while the other, Cu_{a_3} and Fe_{a_3}, comprise the active O_2 binding site. When the active site

* Address correspondence to this author at the Center for Bio-catalysis Science and Technology, Department of Chemistry and Biochemistry, Utah State University, Logan, UT 84322-4630.
‡ Present address: Lawrence Berkeley Laboratory, Berkeley, CA 94720.

0006-2960/89/0428-4461$01.50/0

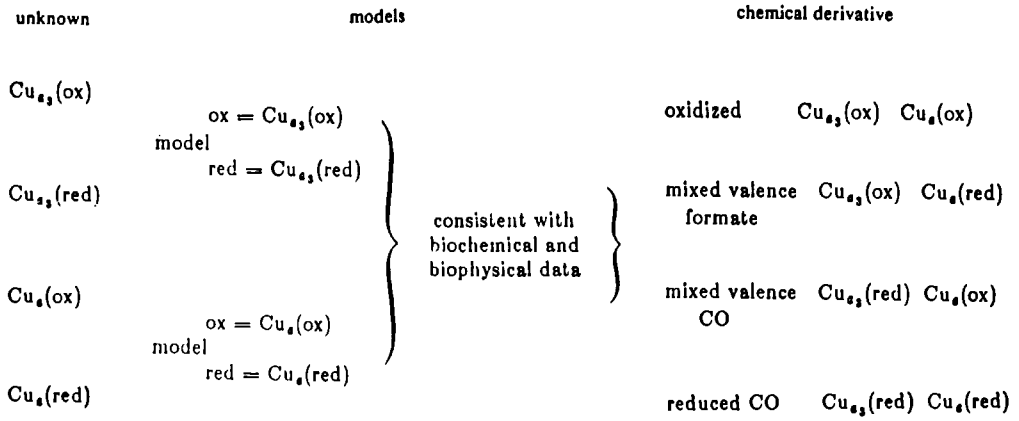

FIGURE 1: Site modeling method as applied to the copper sites of cytochrome *c* oxidase (Powers et al., 1981).

is oxidized, Cu_{a_3} and Fe_{a_3} are in close enough proximity for spin coupling. X-ray absorption spectroscopy can probe the local environment of the metal redox sites but the copper (iron) data are a sum of contributions from both coppers (irons) (Powers et al., 1979, 1981). In this case where two different copper sites with at least two different ligand types are implicated, two analysis approaches are available: model each ligand type with an appropriate model (two atom type procedure) or represent each copper site with an appropriate model (site modeling method). In the latter there are no fitting variables, and the models are further constrained to represent the entire X-ray absorption spectrum of each of the four redox states of the enzyme.

Problems of EXAFS Fitting. It is well documented that EXAFS fitting procedures are difficult when so many ligand contributions (6–12) are possible (Powers et al., 1981; Lee et al., 1981; Peisach et al., 1982; Powers, 1982). The number of degrees of freedom in the filtered X-ray absorption extended fine structure (EXAFS) data of biological samples can at best support fitting procedures containing only two atom types. Each type has the following variable parameters: the number of ligands N at distance r, a Debye–Waller factor $e^{-2\sigma^2k^2}$, and the threshold energy E_0.

Note that SSC do not vary E_0 in their fitting procedures but rather use an arbitrarily chosen point on the absorption edge as a threshold. As is discussed below, an analysis of the eigenvalues and eigenvectors of the Hessian matrix for two-atom fits shows that a linear combination of ΔE_0 values is sometimes a "don't care" parameter having little effect on the fit. In this special case the procedure of SSC may be justifiable. However, the idea of "chemical transferability" of EXAFS phase shifts requires that the relative ΔE_0 be adjusted, whether one is using theoretical phase shifts (Lee & Beni, 1977) or experimental ones (Eisenberger et al., 1978; Eisenberger & Kincaid, 1978; Bunker & Stern, 1983). If this is not done, the ultimate precision in determining distances, which can be better than ±0.01 Å, is lost.

If the ligand contributions are not identical, they are averaged in one of the two atoms types. This produces a host of problems and pitfalls described in detail in the literature, such as the inability to determine uniquely the coordination number of N and S ligands (Peisach et al., 1982) and the difficulty in distinguishing Cu and N ligands in higher coordination shells (Brown et al., 1980). In addition, the amplitudes and phases used in the fitting procedures must be obtained from carefully chosen model compounds or theoretical values (Brown et al., 1980; Stern et al., 1980; Teo, 1981; Peisach et al., 1983; Bunker & Stern, 1983).

Site Modeling. The site modeling method was introduced for the copper and iron contributions of cytochrome *c* oxidase in order to completely avoid these problems and pitfalls of fitting procedures and to incorporate the wealth of spectroscopic and biochemical information available about the redox centers (Powers et al., 1981). The method is shown schematically in Figure 1 for the copper sites and incorporates the following criteria: (1) two model compounds are chosen to be consistent with the spectroscopic and biochemical data such that each model compound in the oxidized and reduced state represents the oxidized and reduced site, respectively; (2) the models in the respective redox states are required to represent the site in four chemical derivatives of the enzyme (fully oxidized, fully reduced with CO, and two mixed-valence states where one site is selectively reduced relative to the other); (3) the *sum* of the model compounds must represent the data of the four derivatives, especially the mixed-valence derivatives, at least as well as any two-atom-type fitting procedure; and (4) the same models must represent all the X-ray absorption data, both edge and EXAFS. SSC misrepresent this method by limiting the comparison to *only* the oxidized-state edge data of the enzyme. There are several combinations of different models that may have biochemical or spectroscopic relevance to the sites that compare favorably to a *single* edge spectrum.

Neither the fitting procedures nor the site modeling method produces mathematically unique results. Both rely on appropriate model compounds suitable to the enzyme sites. The site modeling method, which involves no fitting, is more constrained in that it is required to satisfy equally well the data from four independent chemical derivatives and includes the edge data but attempts to separate the two copper and iron sites. It allows the use of model compounds with complex shell structure and can be used where fitting yields only averaged contributions. When both methods are used properly, fitting and site modeling must yield consistent results.

Analysis Methods

The data were taken from SSC by computer digitizing methods. Comparison of several different digitizing replications showed that they differed by less than the line thickness in the data figures and that these small differences made no differences in the later analysis. In order to ensure that the data were consistent, the EXAFS data of SSC Figure 4g (and Figure 6a) for the K/Bd preparation (measured at 4 K) which had the best signal-to-noise ratio of the reported preparations were Fourier filtered and compared with those of SSC Figure 6c,e. The differences fall within those produced by the respective lengths of the data and by small differences in how the Fourier filter is chosen.

Model Compounds

Model compound data (measured at ca. −150 °C) were treated in the same manner as the enzyme data in length of

the data included, background subtraction, choice of E_0, and Fourier filter. Several models were used to represent the Cu–N(O,C) contribution: tetrakis(imidazole)copper(II) sulfate (Fransson & Lundberg, 1964), copper(II) tetraphenylporphine (CuTPP) (Fleischer et al., 1964), and tetrakis(imidazole)copper(II) and -copper(I) in aqueous solution (Mims & Peisach, 1978) for Cu–N (imidazole), copper(II) diacetate (van Niekerk & Schoening, 1953), and $Cu^{I}(CN)_4$ (Roof et al., 1968). Since C, N, and O differ by only 1 in their atomic number, theoretical EXAFS calculations (Teo, 1981) predict that they are very similar, and they are difficult to distinguish by comparison of EXAFS spectra. Outer-shell contributions from imidazole (C_2, C_4 and N_1, C_5) were represented by that of tetrakis(imidazole)copper(II) and -copper(I) in aqueous solution (CuIm) and tetrakis(imidazole)-copper(II) sulfate and outer-shell carbon (not from imidazole) contributions by the outer shell of copper(II) diacetate. Copper(II) bis(diethyldithiocarbamate) and copper(I) diethyldithiocarbamate (CuDTC) (O'Connor & Maslen, 1966) were used for Cu–S. The Cu–Fe contribution was represented by the Cu–Cu contribution from compounds which are well characterized and similar to binuclear metal sites in proteins such as hemocyanin (Brown et al., 1980), bis(μ-hydroxo)-bis(N,N,N',N'-tetramethylethylenediamine)dicopper(II) bromide (Mitchell et al., 1970), bis(μ-hydroxobis(bipyridyl)dicopper(II) sulfate pentahydrate (Casey et al., 1974), and Cu metal (Thaman et al., 1977). These Cu–Cu contributions can be fit equally well by Cu–Fe, Cu–Cu, and Fe–Cu theoretical amplitudes and phases (Teo, 1981).

Curve Fitting

The fitting procedures employ nonlinear least-squares minimization and are similar to those used by SSC. Although these methods are described by Lee et al. (1981) as applied to X-ray absorption data, it is worthwhile to outline them here.

The goodness of fit is determined by $\sum R^2$, the sum of residuals squared, where $R = (k^3\chi)_{data} - (k^3\chi)_{fit}$. This comparison is appropriate since all data are placed on the same k-space grid for comparison and hence have the same number of points. The goodness of fit value used by SSC, $f' = (\sum R^2/n)^{1/2}/(k^3\chi_{max} - k^3\chi_{min})$, where n is the number of points, was also calculated.

The number of degrees of freedom in the nonlinear least-squares fit is

$$\phi_f = \phi_d - \nu \quad (1)$$

where ϕ_d is the number of degrees of freedom in the filtered data and ν is the number of independent adjustable parameters. An estimate of the number of degrees of freedom in the filtered data can be obtained from

$$\phi_d = 2\Delta r \Delta k/\pi \quad (2)$$

where Δr is the width of the Fourier transform filter window and $\Delta k = k_{max} - k_{min}$ for the data (Lee et al., 1981). This is just the number of Fourier coefficients necessary to describe the filtered data. For the data shown in SSC Figure 6, the first coordination shell was filtered with $\Delta r \approx 1.3$ Å and $\Delta k = 8$ Å^{-1}, so eq 2 gives $\phi_d \approx 6.6$, while the higher shell (Fourier transform peak near 2.7 Å) has a filter width $\Delta r \approx 0.53$ Å and $\phi_d \approx 2.7$. Although the data of Powers et al. (1981) in their Figures 5, 9, and 18 (measured at ca. −100 °C) were treated differently previously, they were treated here in a manner similar to that of SSC for purposes of comparison and have similar ϕ_d. Thus the first-shell data can support fitting procedures having two atom types (eight parameters) while the higher shell can support fitting procedures having at most three variables (one atom type with one of the four required parameters held constant or one atom type phase only procedure). When more variables are included, the fits are overdetermined and the results are meaningless.

Error Estimation. An estimate of the error of each variable parameter can be obtained by varying the parameter away from the best-fit value so that

$$\sum R^2 \rightarrow \sum R^2_{min}(1 + 1/\phi_f) \quad (3)$$

on each side of the minimum while the other parameters are held constant at their minimum values. Therefore, if ϕ_f (eq 1) is large, the estimated error bars will be small, and if ϕ_f is zero or negative, the fit is overdetermined and meaningless. For the first coordination shell data considered here where $\phi_f \approx 0.6$ (N values held constant), a change of $\sum R^2$ on both sides of the minimum (eq 3) by a factor of 2.7 estimates the error and distinguishes two fits having the same number of variable parameters. Similarly, a fit having four variable parameters must decrease the $\sum R^2$ by a factor of 1.4 or more to be judged a better fit. Note that this estimate of the error is strictly true only if the erros are random and therefore represents a best case of minimum error estimate. Systematic errors due to analysis procedures and instrumental functions may also exist, and for this reason, a more conservative estimate has generally been used (Lee et al., 1981; Powers, 1982). The error analyses for variable parameters used by SSC for preparation K/Bd are nearly the same where they chose $\Delta f' = 0.02$ on each side of the minimum except for the number of each ligand type which "were not varied during optimization" and "probably have an error of ± 25%".

The eigenvalues and eigenvectors of the Hessian matrix were examined. The condition number or ratio of the largest to the smallest eigenvalues of the Hessian matrix, $\lambda_{max}/\lambda_{min}$, showed that all the fits reported here were well enough conditioned for the fitting program results to be numerically significant. The eigenvectors corresponding to small eigenvalues represent directions in parameter space along which there is little change in $\sum R^2$ and hence a large uncertainty or error bar for this linear combination of parameters. One is free to redefine the fitting parameters of the problem to be linear combinations of the original parameters, with coefficients determined by the Hessian eigenvectors. This guarantees that there are no correlations among the new parameters, since the eigenvectors are mutually orthogonal. For a *linear* least-squares problem, this new parameterization is independent of the parameter values at the minimum. For a *nonlinear* fit, as in the EXAFS case, the Hessian matrix depends on the parameters in a nonlinear way, such that the new parameterization does not have the same universal meaning as in the linear case.

It is possible to decrease the number of adjustable parameters by holding constant the values of the "don't care" direction linear combinations at or near the best-fit value. These combinations are determined by Hessian eigenvectors at the minimum of the full nonlinear least-squares fit. Holding these parameters constant may not affect the quality of the fit very much and may be useful when ϕ_d is too small to permit a larger number of parameters. For example, in the present case of a two-shell fit to filtered data having $\phi_d \approx 6.6$ (eq 2), a parameter consisting of the sum of the ΔE_0 values for the two models is such a soft or "don't care" parameter, while the difference of the ΔE_0 values is a similar "don't care" parameter with a different error bar. The quality of the fits is such that an almost arbitrary choice of these parameters in a reasonable physical range (with the individual ΔE_0 values varying by as much as ±5 eV) has only a small effect on the value of $\sum R^2$. Such a choice effectively removes two adjustable parameters

from the fit, increasing ϕ_f, and hence possibly improving the estimated error bars. This method is widely used in the analysis of linear least-squares problems (Lawson & Hanson, 1974). This explains the similarity between SSC two-shell fits and those reported here. Such a fortunate correlation of parameters is not always guaranteed to exist, however, and an analysis of a fit with adjustable ΔE_0 is required before the neglect of E_0 as an adjustable parameter is justified.

Site Modeling

As discussed above, model compounds are chosen on the basis of structural, spectroscopic, and other criteria. The structure of the first shell for these models is often complicated enough that the Fourier transform of the EXAFS has the appearance of a "blob" or broad peak, with different distances not fully resolved. Threshold energy E_0 is chosen arbitrarily to be at the middle of the absorption edge. The first-shell broad peaks are then filtered with a wide filter window. The resulting filtered model data then contain only the (unresolved) first-shell contributions. The same filter window is applied to the data to be modeled. Then, with *no* adjustable parameters, the two model data sets are averaged, and the sum of the squares of the residual differences between filtered enzyme data and averaged model data is calculated.

Variations of this technique involve varying the respective E_0's and weighting the model compound data to try to improve $\sum R^2$. In the previously reported work, these adjustments made only statistically insignificant changes in the initial assumptions.

Error Estimation. If the models chosen are exactly correct, the calculated $\sum R^2$ from the above procedure should be dominated by random noise in the data. Since the model compound spectra are relatively noise free compared with those of the enzyme data, it is possible to predict the value of $\sum R^2$ knowing only the noise level of the original data and the number of degrees of freedom in the filtered data. To be statistically significant, a site modeling comparison using different models must produce a $\sum R^2$ larger by a factor of $1 + 1/\phi_d$ (eq 3). In our previous work (Powers et al., 1979, 1981, 1984, 1985; Chance et al., 1983), ϕ_d was about 10 (eq 2). The model compounds chosen did in fact fit the cytochrome oxidase data within the noise, as well as reproducing the edge features in oxidized, reduced, and mixed-valence states, while other plausible models tested resulted in $\sum R^2$ values at least an order of magnitude larger, well outside the statistical limits. As an attempt to include possible systematic effects, the total error estimation was based on a more conservative factor of 2 change in $\sum R^2$.

The major pitfall of this method is the possibility that the deficiencies in the data of one model are compensated for, within the noise, by the data of the other model in each redox state, and thus the model data for each site, although consistent, are not unique. The excellent overall match of the models chosen to the enzyme data (both edge and EXAFS) makes this less likely, but the best way to address this problem is to try different models or to improve the signal-to-noise rato of the data. The previous work still represents the state of the art for cytochrome oxidase data.

Results

Copper Edge. Although SSC claim that the "resting state edge looks very similar to several that have been reported since, including that illustrated in Figure 1", the direct comparison of these edges shows differences that are not within the error as seen in Figure 2a. However, the resting edge of SSC Figure 1 is identical within the error with our results using the Hartzell–Beinert preparation (Naqui et al., 1984) as shown in Figure 2b. The copper edge alone has been shown to be a reliable method of distinguishing differences between purified preparations and membrane-bound preparations and indicates differences in the ligand geometry of the copper atoms (Powers et al., 1987).

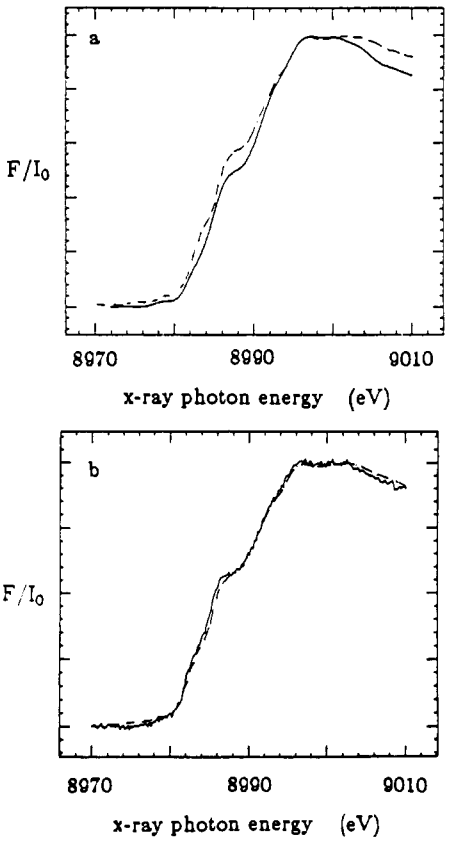

FIGURE 2: Comparison of the X-ray absorption edges of resting cytochrome *c* oxidase: (a) data of Powers et al. (1979) (—) and Scott et al. (1986) (--); (b) data of Naqui et al. (1984) (—) and Scott et al. (1986) (--).

It is clear that the edge data of both groups (Figure 2a) cannot be favorably reproduced by the site modeling method using the same two model compounds since the data are different. It is true, as SSC claim, that either edge alone (not including the other chemical derivatives) can be favorably compared with the combination of several Cu–N, Cu–O, and Cu–S models which have different coordination geometries and ligand contributions. The two models used in the site modeling method for the edge data of the four chemical derivatives (Powers et al., 1979) are also used in the site modeling method with these same derivatives for the copper EXAFS data (Powers et al., 1981).

Copper First-Coordination Shell. Results of the two atom type fitting procedure using a variety of models for Cu–N, Cu–O, and Cu–S contributions are shown in Table I for the first-shell filtered data of SSC Figure 6c for preparation K/Bd. Although the average distance is not model dependent, the change in the Debye–Waller exponent $\Delta\sigma^2$ given as $\sigma^2_{model} - \sigma^2_{enzyme}$, is and has a large error (±45%). The difference in the E_0 for the Cu–N and Cu–S model compounds is ≈8 eV, and the constrained ΔE_0 values used for Tables I and II reflect this fact within the ±5 eV "don't care" range of each. The first set of results reproduce those reported in SSC Table I where the number of N ligands per copper atom N_N was constrained to 3 and number of S ligands per copper atom N_S to 1 for a total of six nitrogen and two sulfur ligands, as shown in Figure 3. Table I also shows solutions found when the

Table I: Results of Two Atom Type Fitting Procedure to the First-Coordination Shell Data of Scott et al. (1986) for Resting Oxidized Cytochrome Oxidase (Preparation K/Bd)[a]

Cu–N			Cu–S			
r_N	N_N	$\Delta\sigma^2$ (×10³)	r_S	N_S	$\Delta\sigma^2$ (×10³)	$\sum R^2$
			N_{total} = 4			
2.00	3	−3.4	2.30	1	6.0	5.0[b]
1.98	2.5	−2.9	2.30	1.5	4.2	4.2
			N_{total} = 3.5			
1.98	2	−0.7	2.30	1.5	4.4	3.1[c]
1.94	1.5	0.3	2.29	2	2.7	4.2
			N_{total} = 3			
1.94	1	4.2	2.29	2	2.3	2.2[c]

[a] r in Å, estimated error ±0.015 Å; N per copper atom and ΔE_0 values held constant during minimization; $\Delta\sigma^2$ in Å², variation ±75% relative to CuTPP for Cu–N and to Cu^{II}DTC for CuS; $\Delta E_{0models}$ = ~8 eV. [b] f' = 0.0335. [c] Best fits.

Table II: Results of Two Atom Type Fitting Procedure for the First-Coordination Shell Data of Powers et al. (1981) for Resting Oxidized Cytochrome Oxidase[a]

Cu–N			Cu–S			
r_N	N_N	$\Delta\sigma^2$ (×10³)	r_S	N_S	$\Delta\sigma^2$ (×10³)	$\sum R^2$
			N_{total} = 4			
2.00	3	−1.6	2.28	1	3.4	10.7
1.98	2.5	−1.3	2.27	1.5	1.2	6.7[b]
1.96	2	−0.4	2.26	2	−0.4	5.4[b]
1.95	1.5	0.7	2.25	2.5	−1.8	5.0[b]
			N_{total} = 3.5			
1.98	2	0.6	2.26	1.5	1.5	5.4[b]
1.95	1.5	1.8	2.25	2	0.0	5.8[b]
			N_{total} = 3			
1.95	1	4.6	2.24	2	0.3	12.5

[a] r in Å, estimated error ±0.015 Å; N per copper atom and ΔE_0 values held constant during minimization; $\Delta\sigma^2$ in Å², variation ±75% relative to CuTPP for Cu–N and to Cu^{II}DTC for CuS; $\Delta E_{0models}$ = ~8 eV. [b] Best fits.

number of each ligand type is changed but held constant during minimization. Results similar to those of Table I were obtained when the data in SSC Figure 4g and also Figure 6a, with $\phi_d \approx 7.4$ (eq 2), were fitted, but the $\sum R^2$ values were roughly a factor of 2 larger. The solutions within each assumed N_{total} of Table I do not have $\sum R^2$ values which differ by a factor of ~1.4 (eq 3) required to differentiate them. Thus the fitting procedure cannot differentiate the number of each ligand type for N_{total} = 4 in these data and does not support "the combination of N_N = 3 and N_S = 1 gave the first fit" which "probably have an error of ±25%" as SSC cladim. However, the solutions of N_N = 1, N_S = 2 (N_{total} = 3) and N_N = 2, N_S = 1.5 (N_{total} = 3.5) are clearly differentiated as best solutions.

This confusion of the number of ligands in the fitting procedure is not restricted to the SSC data. Table II compares the same procedure using the data of Powers et al. (1981) in Figures 5a and 7a, and the behavior is the same. In fact, data synthesized from model compounds having Cu–N and Cu–S contributions exhibit this behavior when fit with the same model compound data from which they were synthesized. The systematics of this fitting confusion has been considered in detail previously for the blue copper protein stellacyanin by Peisach et al. (1982).

It is worthwhile to compare the results of the fitting procedure for the data of both groups. For the data of SSC, best fits are found for $N_N + N_S = N_{total} \leq 3.5$ while the opposite is true for the data of Powers et al., where the fits are worse for $N_{total} < 3.5$. Best fits for the data of both groups have N_S

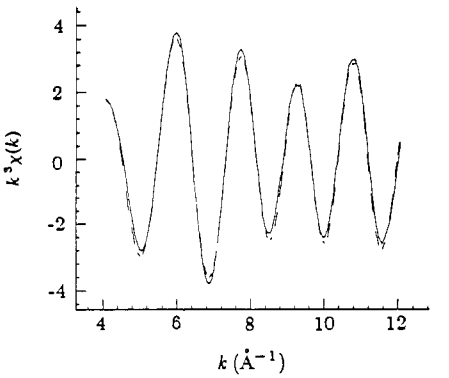

FIGURE 3: First-shell data of Figure 6c of Scott et al. (1986) (—) and first fit of Table I using Cu–N and Cu–S which reproduces the parameters reported therein (– –).

> 1.5. Only one solution has N_N, N_S values common to the data of both groups: N_N = 2, N_S = 1.5 (N_{total} = 3.5). Although the numbers of each ligand type are the same, as are the r_N values, the r_S (±0.015) values are outside the cumulative estimated errors. This estimated error is more conservative than the error obtained by eq 3 and is an attempt to address the possible systematic error. Further, the $\Delta\sigma^2$ values (±75%) for N ligands are also outside the cumulative estimated errors, but this does not take into account the temperature difference between the SSC data and the model compounds used in the analysis. The difference in $\Delta\sigma^2$ for S is in the appropriate direction for this temperature difference, i.e., increases at lower temperature, but that for N is not. When both N values are held constant, $\Delta\sigma^2$ values of the two different atom types are highly correlated. The latter prohibits conclusions about the disorder present in the average distance of each type. The N values of each atom type are also significantly correlated: N_{total} and r for each atom type being the strong parameters.

The site modeling method assumes a particular known model compound for each site based on biochemical and spectroscopic data and incorporates the additional information of other chemical derivatives where the copper atoms can be selectively changed and the edge data in order to determine some information regarding the separate contributions of each copper. The results of the site modeling method for the data of Powers et al. (1981) in Figures 6, 7, 9, and 18 using the same models as used for the edge data [shown in Powers et al. (1979), Figures 16, 17, and 25] are compared with those obtained by the fitting procedure for the mixed-valence derivatives, formate and CO. These selective site reduced derivatives provide the independent tests for the models that were chosen or extracted from the fully oxidized and fully reduced with CO data. In both mixed-valence derivatives, the confusion of the N values in the fitting procedure that was observed for the fully oxidized state is evident. With the site modeling method, $\sum R^2$ is roughly a factor of 2 smaller for the mixed-valence formate derivative when compared with that of the physically reasonable solutions obtained by the fitting procedure (Powers et al., 1981). Since the site modeling method is a sum of data from the two model compounds and involves no fitting or variables, the interference node in this derivative provides a stringent critical differentiation of the models. Comparison with the mixed-valence CO derivative shows that the $\sum R^2$ obtained with the site modeling method is comparable to those obtained by the adjustable parameter fitting procedure. It should be noted that the average of the ligand distances of the models used in the site modeling procedure is within the error of that found by the fitting procedures for all four derivatives as are the respective N values.

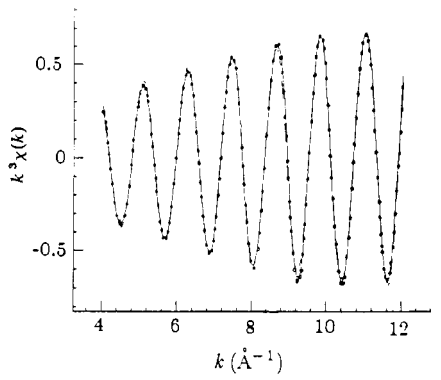

FIGURE 4: Higher shell data of Figure 6e of Scott et al. (1986) (—), first fit of Table III using Cu–C and Cu–Fe which reproduces the parameters reported therein (– –), and second fit of Table III using Cu–Im or C and one-half the parameters of the first fit (···).

This must be true when the two methods are used in the manner in which they are appropriate.

Higher Coordination Shells. Analysis of the outer shell of SSC (Fourier transform peak near 2.7 Å) was also done with the data of SSC Figure 6e, and the results are shown in Table III. The constrained fit of two imidazole contributions [C_2 and C_4 for each] and 0.5 Fe contribution per copper atom reproduces those reported in SSC Table II including the goodness of fit. It is important to note that 1.5 imidazole or outer-shell carbon contributions *alone* with half the number of variable parameters have a comparable $\sum R^2$ as shown in Figure 4. With only $\phi_d \approx 2.7$ (eq 2), this shows the futility of overdetermined fitting.

The data of Powers et al. (1981) have nearly the same ϕ_d for the higher shells as that of SSC, and it was for precisely this reason that "only the phase was used to identify the chemical type of the scattering atom [one type per shell] and the amplitude then tested to be the same as that of the model within small variation of N and/or $\Delta\sigma^2$" in their highest shell data. This method uses only the idea of "chemical transferability" and "standard" phase shifts of EXAFS (Citrin et al., 1980; Teo, 1981; Bunker & Stern, 1983) and fits only two parameters, r and E_0. This procedure applied to the higher shell data of SSC is shown in the other fits of Table III. The fits using imidazole and Cu–C models have $\sum R^2$ roughly one-third as large as fits using Cu–S, Cu–O, Cu–N, or Cu–Fe models and suggest that this shell is largely carbon. Although there may be an iron atom at roughly 3.0 Å in oxidized cytochrome oxidase, this analysis which is not overdetermined does not provide evidence for such.

The results for the higher shell data of Powers et al. (1981) have been described previously. When the N value was fixed at a value determined by other studies, the number of imidazole ligands found by the first shell analysis, 0.5 Fe(Cu), or one sulfur ligand, the results suggested that the second shell contained largely imidazole carbon while the third contains largely iron or copper.

The imidazole contribution (N_1 and C_5 for each) has $\sum R^2$ near that which distinguishes the fits in the third-shell data, and it is quite possible that it contains some imidazole contribution as well. Histidine coordination is suggested in the first-shell analyses, and the outer shells of the enzyme resemble those of model compounds and proteins containing histidine. This possibility can be investigated by subtracting the amount of imidazole contribution found in the first and second shells from the third-shell data, and the remainder fitted as described above to Fe(Cu). Alternatively, a "group" representation can be used in which the outer shells are filtered together as one

Table III: Results of the Fitting Procedure for the Outer-Coordination Shell of Scott et al. (1986) for Resting Oxidized Cytochrome Oxidase

Cu–C (imidazole)				Cu–Fe				
r	N	$\Delta\sigma^2$ (×10³)	ΔE_0	r	N	$\Delta\sigma^2$ (×10³)	ΔE_0	$\sum R^2$
2.91	2.0	8.6		2.94	0.5	2.2		0.23[a,b]
3.02	1.5	5.3						0.27[c]
3.01			−0.5					0.0025[d]
Cu–C (outer shell)				**Cu–Fe**				
r	N	$\Delta\sigma^2$ (×10³)	ΔE_0	r	N	$\Delta\sigma^2$ (×10³)	ΔE_0	$\sum R^2$
2.92	4.0	−1.9		2.96	0.5	−5.0		0.23[a,b]
3.00	3.0	3.3						0.21[c]
3.01			1.5					0.0025[d]
				2.94	0.05	4.0		0.79[c]
				2.95			3.4	0.01[d]
r		N	$\Delta\sigma^2$ (×10³)		ΔE_0		$\sum R^2$	
Cu–C (first shell)								
3.01		0.8	9.2				0.37[c]	
3.02					2.4		0.0016[d]	
Cu–N								
3.01		0.5	9.2				0.71[c]	
3.01					−1.6		0.0081[d]	
Cu–S								
3.11		0.5	5.8				0.61[c]	
3.12					0.9		0.01[d]	

[a] Two atom type fit with N per copper atom and ΔE_0 held constant during minimization; r in Å, estimated error ±0.05 Å; $\Delta\sigma^2$ in Å², variation ±80%. [b] $f' = 0.0355$. [c] Single atom type fit with N per copper atom and ΔE_0 held constant during minimization; r in Å, estimated error ±0.03 Å; $\Delta\sigma^2$ in Å², variation ±80%. [d] Phase shift only fit; r in Å, estimated error ±0.02 Å; ΔE_0 in eV, estimated error ±2 eV; values of $\sum R^2$ for relative comparison only with other phase-only fits.

(Co et al., 1981; Woolery et al., 1984). The imidazole or histidine model compound is similarly filtered, and these data containing the outer shells are treated as a single contribution or atom type in the fitting procedure. Here, $\phi_d \approx 4.4$ (eq 2), and when the number of ligands and Debye–Waller factor of imidazole or histidine are constrained, $\sum R^2$ decreases by more than a factor of 20 when an Fe(Cu) contribution in addition to imidazole is included. However, both methods give the same results within the errors ($r = 3.77 \pm 0.05$ Å) and implicate the presence of a high atomic number atom in the third coordination shell that may be the iron of the binuclear active site.

It is important that although the iron is implicated in the data of Powers et al. (1981) for resting oxidized cytochrome oxidase, the methods used above are by no means conclusive. If this is the case, the binuclear active site contribution must also be observed in the iron data of the resting oxidized state and both the iron and copper data of the mixed-valence formate state where the active site is oxidized. Using the methods described above, Powers et al. (1981) have shown that such contributions are implicated in these data as well but cannot be identified in the mixed-valence CO or reduced CO states where the binuclear active site is reduced. In addition, copper depletion of the oxidized enzyme removed this contribution in the iron data (Powers et al., 1982).

Conclusions

Fitting procedures involving nonlinear least squares have long been known for a multitude of pitfalls and problems, and EXAFS analysis is not exempt. The apparent masking of true S/N by Fourier filtering, the associated loss of degrees of freedom, and the lack of orthogonality of the descriptive variables are relevant considerations. The basic facts remain, however: the number of variables in the fit must be less than

the number of independent degrees of freedom in the data; the larger the difference, the more easily solutions can be differentiated within the limitations of the signal to noise of the data. The fit must be numerically well conditioned: calculations done by the fitting program must lie within the dynamic range of the machine. Orthogonal directions in parameter space (eigenvectors of the Hessian matrix) along which there is little change in the quality of the fit, i.e., linear combinations of parameters having a large estimated error, can be dropped from the fit and hence reduce the number of fitting parameters for a given number of degrees of freedom in the data. An examination of the eigenvalues and eigenvectors of the Hessian matrix is especially useful for these purposes.

For EXAFS data, other information in addition to good model compound data is often needed to guide the analysis and to distinguish physically reasonable solutions. For the first-shell analysis of many proteins and model compounds the number of nitrogen, carbon, oxygen, or sulfur ligands could not be distinguished by fitting procedures alone. Additional information from biochemical and spectroscopic studies was needed before information useful to this problem could be obtained. When such information is not available, the site modeling method is reduced to the trial and error of guessing possible model compounds. Neither fitting procedures nor the site modeling method gives unique solutions, but both approaches must be consistent when applied in the appropriate manner.

The fitting procedure shows differences in the data of SSC and Powers et al. (1981) for oxidized cytochrome oxidase that are outside the cumulative estimated errors and are independent of any assumptions. The best solutions for the SSC data have $N_{total} \leq 3.5$ while those of Powers et al. (1981) are best for $N_{total} \geq 3.5$. Only one N_S solution is common to the data of both groups. However, the r_S values for both data are not within the cumulative estimated errors. SSC find only one higher shell for which they over interpret to contain four carbon and 0.5 iron contributions, but only the carbon contribution can be justified when the fit is not overdetermined. The data of Powers et al. (1981) also show carbon in the second shell. A higher shell is observed in the data of Powers et al. (1981), which when analyzed together with the second shell may contain imidazole contributions and a higher atomic number contribution. Inclusion of the iron EXAFS data of the resting oxidized state together with both the iron and copper data of the mixed-valence and reduced CO states (Powers et al., 1981) as well as the iron data of the copper-deleted oxidized enzyme (Powers et al., 1982) suggests that this higher atomic number contribution could be the binuclear active site. One can identify scattering from iron in the copper data and from copper in the iron data. This contribution is only slightly altered in the pulsed (Chance et al., 1983) and pulsed peroxide states which are also fully oxidized (Powers & Chance, 1984, 1985; Powers et al., 1988).

Although SSC claim that the site modeling method does not support the existence of sulfur ligands, lending credence to their model which placed no sulfur ligands on the copper of the active site in the oxidized state, their recent results for Cu_a-depleted, *p*-(hydroxymercuri)benzoate-modified enzyme show a sulfur ligand on the active site copper (Li et al., 1987), in agreement with the results of the site modeling method. Unfortunately, the analysis of these data also suffers from the same multivariable analysis of the number of each type of ligand as was illustrated above. It is unlikely that a sulfur ligand is methionine since methionine is not known to bind copper except under harsh and unphysiological conditions (Sigel et al., 1980). This amino acid is known to be associated with the blue copper proteins; however, the distances are probably too long ($r \gtrsim 2.7$ Å) (Coleman et al., 1978; Adman et al., 1978) for bonding.

Although the technique of X-ray absorption spectroscopy and the appropriate analysis procedures provide a very powerful tool for the study of the local structure around the active sites of biological molecules which cannot be crystallized or which have spectroscopic signatures that are difficult to interpret, these procedures must be used with care and a clear understanding of the procedures employed if meaningful results are to be gained.

Registry No. Cytochrome *c* oxidase, 9001-16-5.

References

Adman, E., Stenkamp, R., Sieker, L., & Jensen, L. (1978) *J. Mol. Biol. 123*, 35–48.

Brown, J., Powers, L., Kincaid, B., Larrabee, J., & Spiro, T. (1980) *J. Am. Chem. Soc. 102*, 4210–4216.

Bunker, B., & Stern, E. (1983) *Phys. Rev. B27*, 1017–1027.

Casey, A., Hoskins, B., & Whillans, F. (1974) *Chem. Commun.*, 904–905.

Chance, B., Powers, L., Ching, Y., & Kumar, C. (1983) *Biophys. J. 44*, 353–363.

Citrin, P., Eisenberger, P., & Kincaid, B. (1976) *Phys. Rev. Lett. 36*, 1346–1349.

Co, M., Scott, R., & Hodgson, K. (1982) *J. Am. Chem. Soc. 103*, 986–989.

Coleman, P., Freeman, H., Guss, J., Murata, M., Norris, V., Ramshaw, J., & Venkatappa, M. (1978) *Nature (London) 272*, 319–324.

Eisenberger, P., & Kincaid, B. M. (1978) *Science 200*, 1441–1446.

Eisenberger, P., & Lengler, B. (1980) *Phys. Rev. B22*, 3551–3562.

Eisenberger, P., Shulman, R. G., Kincaid, B. M., Brown, G. S., & Ogawa, S. (1978) *Nature (London) 274*, 30–34.

Fleischer, E., Miller, C., & Webb, L. (1964) *J. Am. Chem. Soc. 86*, 2342–2347.

Fransson, G., & Lundberg, B. (1972) *Acta Chem. Scand. 26*, 3969–3976.

Lawson & Hanson (1974) *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ, and EISPACK Library Subroutine MINFIT, Applied Mathematics Division, Argonne National Laboratory, Argonne, IL.

Lee, P., Citrin, P., Eisenberger, P., & Kincaid, B. (1981) *Rev. Mod. Phys. 53*, 769–806.

Li, P., Chan, S., Sullivan, R., & Scott, R. (1987) *Biochemistry 26*, 2091–2095.

Mims, W., & Peisach, J. (1978) *J. Chem. Phys. 69*, 4921–4929.

Mitchell, T., Bernard, W., & Wasson, J. (1970) *Acta Crystallogr., B26*, 2096–2101.

Naqui, A., Kumar, C., Ching, Y., Powers, L., & Chance, B. (1984) *Biochemistry 23*, 6222–6227.

O'Connor, B., & Maslen, E. (1966) *Acta Crystallogr. 21*, 828–830.

Peisach, J., Powers, L., Blumberg, W., & Chance, B. (1982) *Biophys. J. 38*, 277–285.

Powers, L. (1982) *Biochim. Biophys. Acta 683*, 1–38.

Powers, L., & Chance, B. (1984) in *EXAFS and Near Edge Structure* (Hodgson, K., Hedman, B., & Penner-Hahn, J., Eds.) Vol. III, pp 117–126, Springer-Verlag, Berlin.

Powers, L., & Chance, B. (1985) *J. Inorg. Biochem. 23*, 207–217.

Powers, L., Blumberg, W. E., Chance, B., Barlow, C., Leigh, J., Jr., Smith, J., Yonetani, T., Vik, S., & Peisach, J. (1979) *Biochim. Biophys. Acta 547*, 520-538.
Powers, L., Chance, B., Ching, Y., & Angiolillo, P. (1981) *Biophys. J. 34*, 465-498.
Powers, L., Chance, B., Ching, Y., Muhoberac, B., Weintraub, S., & Wharton, D. (1982) *FEBS Lett. 138*, 245-248.
Powers, L., Chance, B., Ching, Y., & Lee, C. (1987) *J. Biol. Chem. 262*, 3160-3164.
Powers, L., Naqui, A., Kumar, C., Ching, Y., & Chance, B. (1988) *J. Biol. Chem. 263*, 7159-7165.
Roof, R., Jr., Larson, A., & Cromer, D. (1968) *Acta Crystallogr. B24*, 269-276.
Scott, R., Schwartz, J., & Cramer, S. (1986) *Biochemistry 25*, 5546-5555.
Sigel, H., Scheller, K., Rheinberger, V., & Fischer, B. (1980) *J. Chem. Soc., Dalton Trans.*, 1022-1028.
Teo, B. (1981) in *EXAFS Spectroscopy: Techniques and Applications* (Teo, B., & Joy, D., Eds.) pp 13-58, Plenum Press, New York.
Thaman, T., Loehr, J., & Loehr, T. (1977) *J. Am. Chem. Soc. 99*, 4187-4193.
van Niekerk, J., & Schoening, F. (1953) *Acta Crystallogr. 6*, 227-232.
Woolery, G., Powers, L., Winkler, M., Solomon, E., & Spiro, T. (1984) *J. Am. Chem. Soc. 106*, 86-92.

Presence of the HNK-1 Epitope on Poly(*N*-acetyllactosaminyl) Oligosaccharides and Identification of Multiple Core Proteins in the Chondroitin Sulfate Proteoglycans of Brain†

D. C. Gowda, R. U. Margolis,* and R. K. Margolis

Department of Pharmacology, New York University Medical Center, New York, New York 10016, and Department of Pharmacology, State University of New York, Health Science Center at Brooklyn, Brooklyn, New York 11203

Received October 28, 1988; Revised Manuscript Received February 1, 1989

ABSTRACT: The chondroitin sulfate proteoglycans of brain contain several core proteins bearing HNK-1 antibody epitopes. Endo-β-galactosidase treatment resulted in the almost complete disappearance of HNK-1 staining of proteoglycan immunoblots, indicating that a significant portion of the 3-sulfated sugar residues recognized by this antibody are present on poly(*N*-acetyllactosaminyl) oligosaccharides. However, after treatment with chondroitinase ABC followed by endo-β-galactosidase, several proteoglycan species showed HNK-1 reactivity, presumably due to the presence of this epitope on other oligosaccharides which are both resistant to endo-β-galactosidase and inaccessible to the antibody in the native proteoglycan. Immunostaining of the endo-β-galactosidase degradation products after separation by thin-layer chromatography demonstrated that HNK-1 reactivity was confined to a minor population of large oligosaccharides. Only a relatively small portion of the native chondroitin sulfate proteoglycans of brain enter a 6-12% SDS-polyacrylamide gel. However, after treatment of the proteoglycans with chondroitinase ABC (or chondroitinase and endo-β-galactosidase) in the presence of protease inhibitors, seven bands with molecular sizes ranging from 80 to 200 kDa appear in Coomassie Blue stained gels, and two additional bands with molecular sizes of 67 and 350-400 kDa are apparent in fluorographs of sodium [^{35}S]sulfate labeled proteoglycans. Most of these components probably represent individual proteoglycan species rather than different degrees of nonchondroitin sulfate/keratan sulfate glycosylation of a single protein core, since [^{35}S]methionine-labeled proteins of comparable molecular size were synthesized by an in vitro translation system. These findings suggest that chondroitin sulfate proteoglycans which differ in molecular size and composition may be specific to particular cell types in brain.

The chondroitin sulfate proteoglycans of brain range in molecular size from approximately 260 to 325 kDa, on the basis of their gel filtration behavior under dissociative conditions (Krusius et al., 1987). They are mostly soluble in a phosphate-buffered saline extract and account for less than 1% of the soluble brain protein (Kiang et al., 1981). Immunoelectron microscopic studies have demonstrated that the chondroitin sulfate proteoglycans are present in the extracellular space of early postnatal brain, after which period they progressively assume an intracellular (cytoplasmic) localization in neurons and astrocytes (Aquino et al., 1984a,b). Although biochemical assays have demonstrated only a limited degree of aggregation with hyaluronic acid, recent comparative studies on the localization of hyaluronic acid, hyaluronic acid binding region, link protein, and chondroitin sulfate proteoglycans in developing rat cerebellum suggest that much of the chondroitin sulfate proteoglycan of brain may occur in the form of aggregates with hyaluronic acid (and link protein) in situ (Ripellino et al., 1988, 1989).

The chondroitin sulfate proteoglycans contain an average of 56% protein, 24% glycosaminoglycans (predominantly chondroitin 4-sulfate, accompanied by smaller amounts of chondroitin 6-sulfate and keratan sulfate), and 20% N- and O-glycosidic glycoprotein oligosaccharides (Kiang et al., 1981;

† This work was supported by grants from the National Institutes of Health and the National Institute of Mental Health (NS-09348, NS-13876, and MH-00129).

* Address correspondence to this author at the Department of Pharmacology, New York University Medical Center.

0006-2960/89/0428-4468$01.50/0