Biophysical Chemistry 4 (1976) 123–130

PREDICTION OF THE SECONDARY AND TERTIARY STRUCTURE OF PLASTOCYANIN

Donald G. WALLACE*
Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road N.W., Washington, D.C. 20015, USA

Received 18 July 1975

The amino acid sequences of nine plastocyanins were examined using four published methods for the prediction of secondary structure in proteins. The results of the four methods were combined in such a way as to maximize agreement, and the position of α helices, β sheets, and β turns in plastocyanin was predicted. From this result and other information, such as the position of conserved residues and the requirements for coordination of copper, a preliminary model for the mainchain folding of the molecule was presented.

1. Introduction

Plastocyanins are deep blue copper proteins involved in photosynthetic electron transport. Information about the secondary and tertiary structures of these molecules is of considerable interest since it may shed light on the nature of the coordination of the copper, as well as on the mechanism of electron transfer in this and other copper proteins.

Several methods have been developed for the prediction of secondary structures in proteins. Three of the most convenient are the methods of Chou and Fasman [1], Lim [2], and Kabat and Wu [3]. A fourth method, that of Wu and Kabat [4], while requiring the use of a computer, is simple in principle and also provides values for the ϕ, ψ angles at each position. Since the amino acid sequences of a number of higher plant and algal plastocyanins have recently become available [5–11], it was decided to apply the above-mentioned four methods to these sequences.

This paper presents a prediction of the secondary structure of plastocyanin and a tentative model for the main-chain folding of the molecule. In the not too distant future it will be possible to compare these results with those obtained by X-ray analysis of plastocyanin. Freeman and his colleagues at the University of Sydney have already begun such a study (Freeman, H.C., personal communication). In addition, the solution of the X-ray structure of bacterial azurin is nearing completion (Jensen, L.H., personal communication); and since this molecule shares some sequence homology with plastocyanin [12,13], it may well have a similar tertiary structure.

* Present address: Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138. All correspondence to be sent to this address.

2. Materials and methods

The complete amino acid sequences of plastocyanins from the green alga *Chlorella fusca* [5]; French bean, *Phaseolus vulgaris* [6]; broadbean, *Vicia faba* [7]; potato, *Solanum tuberosum* [7]; elder, *Sambucus nigra* [7]; vegetable marrow, *Cucurbita pepo medullosa* [8]; spinach, *Spinacia oleracea* [9]; the shepherd' purse *Capsella Bursa-pastoris* [10] were examined. In addition, an incomplete sequence for the green alga *Scenedesmus obliquus*, lacking residues 43–54 [11], was employed. The sequence numbering used here designates the *Chlorella* N-terminal ASP as 1 and the C-terminal residue GLN as 100, with the presumed copper-coordinating CYS as 85.

For the methods of Chou and Fasman, Lim and Kabat and Wu, each sequence was examined one at a time. Then, the best final prediction for each of the

Table 1
Prediction of the secondary structure of plastocyanin by four methods

Method[a]:	Fasman	Lim	Kabat and Wu	Wu and Kabat	Final prediction
Structures[b] predicted by residue number:	2–6 β (7–10 βturn) 9–12 βturn	2–6 βint	2–6 β	2–6 β 7–10 βturn	2–6 β (7–10 βturn)
	12–16 β 16–19 βturn	13–15 βint	11–16 β or α (21–24 β)	14–16 β or α 16–19 βturn	12–16 β 16–19 βturn
	27–31 β 39–43 β 41–47 α 47–50 βturn	26–31 βint 38–43 βint	27–31 β 39–43 α or β	27–32 β 39–42 β or α 41–47 α 48–51 βturn	27–31 β 39–43 β (α) 47–50 βturn
			49–55 α 51–55 β	50–56 α	51–56 α
	58–67 α	63–71 α	58–65 α	59–65 α	59–65 α
	70–75 β 79–82 βturn 81–84 βturn 86–89 βturn	71–75 βsurf 83–85 βint	(73–75 β) 87–90 α	70–74 β 79–82 βturn 80–83 β 86–89 βturn 88–93 α	70–75 β (79–82 βturn) 86–89 βturn
	93–99 β		91–97 β or α	96–100 β	93–99 β

[a] Predictions were carried out separately on nine plastocyanin sequences by the following methods: Fasman [1], Lim [2], Kabat and Wu [3], and Wu and Kabat [4].

[b] Secondary structures for plastocyanin residues 2–100 are tabulated for each method, and a final best prediction is given. Structures in parentheses are for predictions which are weaker. The method of Lim [2] distinguishes between internal (int) and surface (surf) β sheets. The method of Kabat and Wu assigns segments in which structures are *permitted*; it does not require that they be there. Similarily, the method of Wu and Kabat merely confirms whether the structures predicted in any of the other three methods are possible; it neither makes new predictions nor necessarily prohibits others. Terms used are: α, α helix, including 3_{10} helix; β, β sheet; β turn, the reverse turn, as defined by Venkatachalam [15].

three methods was arrived at by maximizing agreement among the nine sequences. This approach assumes that proteins which are closely related evolutionarily will have the same secondary and tertiary structure. Therefore, the predictions for each individual species will tend toward the same actual structure for the protein.

As an example, let us consider the N-terminal region of the plastocyanin molecule. When the method of Fasman was applied, $\langle P_\beta \rangle$ was slightly less than $\langle P_\alpha \rangle$ for elder, potato, broadbean, French beam, *Scenedesmus*, spinach, and marrow sequences (marrow: $\langle P_\alpha \rangle = 1.15$, $\langle P_\beta \rangle = 1.13$). In two sequences, from shepherd's purse and *Chlorella*, $\langle P_\beta \rangle$ was slightly greater than $\langle P_\alpha \rangle$. However, this segment was not predicted as an α helix because it is only five residues long (GLY 7 is a strong helix breaker; see table 1 for the sequence of potato plastocyanin). Six residues are the minimum specified by Fasman for stable helix formation. The prediction for all plastocyanins is a β sheet at residues 2–6. (See table 1).

Turning to the method of Lim, again the nine sequences were examined at the N-terminal region. A β sheet at residues 2–6 was predicted for all nine. No α helix was predicted in any of the sequences because VAL 4 (or ILE 4 in elder) is an H_L residue which cannot form a hydrophobic pair (1–5), (1–4), or hydrophobic triplets (1–2–5) or (1–4–5). This violates rule 1b for helix formation, according to Lim. In the method of Kabat and Wu for the N-terminal segment, an α helix was permitted in elder, and β sheets were permitted in *Chlorella*, potato, *Scenedesmus*, and shepherd's purse. The choice for this method was 2–6 β sheet. In general, assignment of a structure was not made unless it was predicted by several sequences.

The method of Wu and Kabat [4] was carried out with the aid of an IBM 1130 Computer at the Carnegie Institution of Washington, using the set of ϕ, ψ angles for eleven proteins deposited by these authors in the

Table 2
Prediction of ϕ, ψ angles for plastocyanin by the method of Wu and Kabat[a)]

Sequence of potato plastocyanin[b)]	Peptide	Choices of angles[c)]					
		First		Second		Third	
LEU							
ASP	2–3–4	$-63, -47_{I}$	(13)	$-40, -60_{I}$	(11)	$-150,+160_{I}$	(3)*
VAL	3–4–5	$-50, -49_{I}$	(13)	$-132,+141_{I}$	(9)*		
LEU	4–5–6	$-54, -44_{I}$	(5)	$-122,+136_{II}$	(5)*	$-95,+134_{I}$	(4)
LEU	5–6–7	$-58, -51_{I}$	(4)	$-71, -49_{I}$	(4)	$-124,+146_{I}$	(2)*
GLY	6–7–8	$-52, -51_{I}$	(6)	$-68, -53_{II}$	(4)*	$-98,+154_{I}$	(4)
GLY	d)→7–8–9	$-51, -58_{I}$	(12)*	$-93,+151_{II}$	(7)		
ASP	→8–9–10	$-47, -52_{II}$	(8)*	$-81, +10_{I}$	(4)		
ASP	→9–10–11	$-102, +5_{I}$	(8)*	$-67, -48_{I}$	(6)	$-37,+135_{I}$	(4)
GLY	→10–11–12	$-51, -55_{II}$	(5)*	$+91, -12_{II}$	(4)		
SER	11–12–13	$-147,+142_{I}$	(6)*	$-49, -50_{II}$	(5)	$-90,+133_{III}$	(4)
LEU	12–13–14	$-50, -48_{I}$	(10)	$-114, -5_{I}$	(9)*		
ALA	13–14–15	$-59, -47_{II}$	(3)	$-140,+134_{III}$	(2)*		
PHE	14–15–16	$-41, -61_{II}$	(8)	$-108,+103_{I}$	(5)*	$-70, -40_{II}$	(4)
ILE	15–16–17	$-108,+106_{III}$	(5)*	$-51, -51_{III}$	(2)	$-104,+141_{III}$	(2)
PRO	–16–17–18	$-55,+131_{I}$	(17)*	$-53, -44_{I}$	(17)		
GLY	→17–18–19	$-92, -6_{I}$	(12)*	$-39, -60_{II}$	(7)		
ASN	18–19–20	$-49, -51_{I}$	(11)*	$-108, -10_{I}$	(8)		
PHE	19–20–21	$-46, -60_{II}$	(7)*	$-110,+135_{II}$	(6)	$-69, -32_{I}$	(5)
SER	20–21–22	$-57, -51_{II}$	(5)*	$-82,+129_{III}$	(2)		
VAL	21–22–23	$-106,+133_{I}$	(11)	$-52, -51_{I}$	(11)*		
SER	22–23–24	$-49, -51_{I}$	(15)*	$-69, -40_{II}$	(13)	$-152,-180_{I}$	(9)
ALA	23–24–25	$-46, -52_{I}$	(15)*	$-104,+155_{I}$	(11)		
GLY	24–25–26	$-49, -54_{I}$	(8)	$+75, +40_{I}$	(8)*		
GLU	→25–26–27	$-48, -54_{II}$	(5)*	$-91, +18_{I}$	(2)	$-75, -34_{III}$	(2)
LYS	→26–27–28	$-81, -16_{I}$	(7)	$-139,+141_{I}$	(5)*	$-48, -52_{II}$	(5)
ILE	27–28–29	$-45, -57_{II}$	(6)	$-130,+140_{II}$	(5)*	$-80, -20_{I}$	(4)
THR	28–29–30	$-41, -59_{II}$	(2)	$-124,+148_{III}$	(2)*		
PHE	29–30–31	$-93, +94_{I}$	(5)*	$-49, -51_{II}$	(4)	$-85,+148_{II}$	(4)
LYS	30–31–32	$-77,+139_{I}$	(4)*	$-157,+145_{II}$	(3)	$-50, -55_{II}$	(2)
ASN	→31–32–33	$-90, +27_{I}$	(4)	$-99,+137_{II}$	(4)*	$-49, -58_{III}$	(2)
ASN	→32–33–34	$-88, +1_{III}$	(1)*	$-70, -41_{III}$	(1)		
ALA	→33–34–35	$-67, -29_{II}$	(2)	$-88,+154_{III}$	(1)*	$-44, -67_{III}$	(1)
GLY	34–35–36	$-100, -34_{I}$	(3)	$+65, +36_{I}$	(3)*		
PHE	35–36–37	$-123, +49_{I}$	(3)*	$-95,+163_{III}$	(1)		
PRO	36–37–38	$-71, -49_{I}$	(3)*	$-37, -42_{III}$	(1)		
HIS	37–38–39	$-49, -54_{III}$	(1)	$-64, -33_{III}$	(1)*		
ASN	38–39–40	$-40, -61_{III}$	(1)	$+50, +42_{II}$	(2)*	$-135,+113_{III}$	(1)
VAL	39–40–41	$-57, -46_{II}$	(4)	$-130,+149_{II}$	(3)*		
VAL	40–41–42	$-126,+140_{II}$	(3)*	$-94,+124_{III}$	(2)	$-50, -63_{III}$	(1)
PHE	41–42–43	$-96,+102_{II}$	(2)*	$-85,+135_{III}$	(1)	$-52, -46_{III}$	(1)
ASP	→42–43–44	$-57, -51_{II}$	(3)*	$-75, -18_{II}$	(2)		

Table 2 (continued)

Sequence	Peptide	First		Second		Third	
GLU	→43–44–45	$-68, -27_{II}$	(3)*	$-41, -52_{III}$	(1)	$-86,+135_{III}$	(1)
ASP	44–45–46	$-51, -54_{II}$	(4)*	$-60, -33_{III}$	(2)		
GLU	45–46–47	$-46, -53_{I}$	(6)*	$-150,+141_{III}$	(1)	$-74,+135_{III}$	(1)
ILE	46–47–48	$-58, -56_{III}$	(1)*	$-91,+135_{III}$	(1)		
PRO	47–48–49	$-45, -48_{I}$	(7)*	$-68,+151_{II}$	(4)		
ALA	→48–49–50	$-46, -51_{I}$	(6)	$-123,+152_{I}$	(8)*	$-62, -39_{II}$	(2)
GLY	→49–50–51	$-49, -53_{I}$	(7)	$-64, -40_{I}$	(5)*	$-110, -10_{II}$	(4)
VAL	→50–51–52	$-48, -58_{II}$	(4)	$-129,+161_{I}$	(4)	$-65, -42_{I}$	(4)*
ASP	51–52–53	$-67, -53_{II}$	(4)*	$-49, -52_{II}$	(5)	$-90,-169_{I}$	(3)
ALA	52–53–54	$-45, -60_{II}$	(11)*	$-64, -54_{II}$	(10)		
SER	→53–54–55	$-64, -43_{I}$	(14)*	$-46, -50_{I}$	(13)	$-129,+147_{II}$	(8)
LYS	→54–55–56	$-70, -43_{I}$	(10)*	$-46, -56_{I}$	(10)	$-79,+133_{II}$	(5)
ILE	→55–56–57	$-47, -50_{I}$	(4)*	$-94, -6_{I}$	(3)	$-89,+151_{II}$	(3)
SER	→56–57–58	$-122,+130_{II}$	(2)*	$-51, -22_{III}$	(1)		
MET	→57–58–59	$-91,+145_{I}$	(8)*	$-100, +1_{I}$	(4)		
ALA	→58–59–60	$-60, -54_{II}$	(6)*	$-80, -45_{III}$	(2)		
GLU	→59–60–61	$-45, -53_{I}$	(12)*	$-89, -6_{I}$	(10)		
GLU	→60–61–62	$-68, -37_{II}$	(9)*	$-45, -56_{I}$	(5)		
ASP	→61–62–63	$-64, -42_{I}$	(9)*	$-50, -53_{I}$	(8)	$-126, +37_{I}$	(3)
LEU	62–63–64	$-70, -47_{I}$	(6)*	$-47, -52_{II}$	(5)	$-123,+140_{II}$	(3)
LEU	63–64–65	$-50, -45_{II}$	(6)*	$-64, -40_{II}$	(3)		
ASN	64–65–66	$-53, -55_{II}$	(3)*	$-71, -46_{II}$	(3)		
ALA	→65–66–67	$-65, -38_{II}$	(3)	$-91,+132_{II}$	(4)*		
ALA	→66–67–68	$-50, +99_{I}$	(6)*	$-88, -33_{I}$	(3)	$-135,+146_{I}$	(3)
GLY	67–68–69	$-45, -53_{I}$	(4)*	$+71, +47_{I}$	(4)	$+83, -66_{II}$	(3)
GLU	→68–69–70	$-107,+153_{II}$	(6)*	$-52, -52_{II}$	(4)	$-64, -41_{III}$	(3)
THR	→69–70–71	$-88, -11_{I}$	(5)	$-56,+121_{II}$	(4)*		
TYR	→70–71–72	$-96,+103_{II}$	(8)*	$-63, -33_{II}$	(8)	$-104,+139_{II}$	(6)
SER	→71–72–73	$-39, -54_{II}$	(9)	$-67, -37_{II}$	(5)	$-136,+145_{II}$	(4)*
VAL	72–73–74	$-49, -52_{II}$	(15)	$-111,+142_{I}$	(14)*	$-64, -46_{II}$	(6)
THR	→73–74–75	$-105,+142_{I}$	(11)*	$-65, -40_{I}$	(10)	$-49, -60_{I}$	(6)
LEU	→74–75–76	$-84,+114_{I}$	(7)*	$-69, -37_{I}$	(7)	$-124, +43_{I}$	(4)
SER	→75–76–77	$-53, -65_{I}$	(11)*	$-66, -36_{II}$	(8)	$-56,+145_{I}$	(5)
GLU	→76–77–78	$-48, -51_{II}$	(10)*	$-103, +1_{I}$	(6)	$-135,+150_{II}$	(6)
LYS	77–78–79	$-48, -60_{I}$	(11)*	$-67, -50_{II}$	(9)	$-104,+159_{III}$	(5)
GLY	78–79–80	$-48, -48_{I}$	(11)	$-66, -49_{I}$	(9)	$+76, +34_{I}$	(9)*
THR	→79–80–81	$-42, -56_{I}$	(5)	$-108,+143_{III}$	(4)	$-64, -35_{III}$	(4)*
TYR	→80–81–82	$-112,+144_{I}$	(8)	$-65, -37_{II}$	(7)*	$-51, -51_{I}$	(6)
THR	→81–82–83	$-49, -54_{II}$	(7)*	$-104, +2_{I}$	(5)	$-122, +97_{I}$	(5)
PHE	82–83–84	$-71, -18_{I}$	(5)*	$-71,+118_{III}$	(4)		
TYR	83–84–85	$-49, -62_{III}$	(1)*				
CYS	84–85–86	$-40, -58_{II}$	(4)*	$-113,+142_{III}$	(3)		
ALA	85–86–87	$-102,+129_{II}$	(4)*	$-153,+159_{III}$	(1)		
PRO	→86–87–88	$-74, -32_{I}$	(6)*				
HIS	→87–88–89	$-104, -3_{III}$	(1)	$-54, -52_{III}$	(1)*		

Table 2 (continued)

Sequence	Peptide	First		Second		Third	
GLN	88–89–90	$-46, -55_{II}$	(3)*	$-60, -29_{III}$	(1)		
GLY	89–90–91	$-59, -45_{I}$	(4)*	$-53, -55_{II}$	(3)		
ALA	90–91–92	$-47, -52_{II}$	(2)*	$-61, -30_{II}$	(2)		
GLY	91–92–93	$-44, -56_{III}$	(1)*	$+68,-142_{III}$	(1)	$-121,+149_{III}$	(1)
MET	92–93–94	$-45, -60_{III}$	(2)	$-69, -9_{III}$	(2)*		
VAL	→93–94–95	$-52, -63_{III}$	(2)*	$-63, -43_{III}$	(2)		
GLY	→94–95–96	$-62, -40_{I}$	(5)*	$-50, -55_{II}$	(4)		
LYS	→95–96–97	$-58, -54_{II}$	(8)	$-120,+153_{II}$	(6)*	$-69, -33_{II}$	(4)
VAL	→96–97–98	$-50, -56_{II}$	(6)	$-64, -43_{II}$	(6)	$-116,+139_{II}$	(5)*
THR	97–98–99	$-49, -59_{I}$	(7)	$-137,+148_{II}$	(6)*		
VAL	98–99–100	$-48, -49_{I}$	(7)	$-120,+140_{II}$	(5)*	$-95, -37_{III}$	(2)
ASN							

a) Wu and Kabat [4].

b) The amino acid residue is given for the nth position of each tripeptide: $n-1$, n, $n+1$. The numbering is from 2 (N-terminal to 100 (C-terminal).

c) For each tripeptide the computer searched the X-ray data from eleven proteins and tabulated ϕ, ψ angles for each tripeptide for each of the nine plastocyanin sequences. The angles which occurred most frequently at each position are presented. The subscript Roman numerals I, II, and III are as defined by Wu and Kabat [4] and represent the highest criterion observed for that angle pair. The numbers in parentheses are the sums of weighted frequencies of occurrence, weighted as follows: criterion I, 3; criterion II, 2; criterion III, 1; criterion IV, 0. The choices labelled with an asterisk are the ones used to construct the potato plastocyanin model.

d) Tripeptides which possess angle-pair choices within the range of β turns [15]. $C_{\alpha 2}$ and $C_{\alpha 3}$ residues of the possible turns are marked. See text for details.

Protein Data Bank, Brookhaven National Laboratory [14]. Tripeptides from each plastocyanin sequence were examined separately at each position; the best ϕ, ψ angles at each position were then chosen from these results. (see table 2). At first it was attempted to deduce criteria by which secondary structures could be predicted. Tests were made on the similar set of angle pairs compiled by Wu and Kabat for cytochrome *c* [4]. Preliminary results, using the frequency of occurrence as the basis for choosing a structure, were not encouraging. Examination of table 1 reveals many angle pairs in the α helical domain [3], near $\phi = -57$, $\psi = -55$. This is understandable because several of the eleven proteins from which ϕ, ψ angles were taken are helixrich. Therefore, predictions from this data may be biased in favor of helical versus β sheet structures. For the purposes of prediction of plastocyanin secondary structure, it was decided to use the ϕ, ψ angles for the nine plastocyanins as a confirmatory source of information, rather than as a primary predictive method, per se. For example, in the segment 2–6 considered above, the other three methods predicted a β sheet. In table 2, one can readily pick out angle-pair choices with the β sheet configuration in 2–6 ($\phi = -70$ to -165, $\psi = +80$ to $+170$). However, these choices are second or third; a first choice prediction would be helical. In practice, then, the ϕ, ψ angles were only tested for possible consistency with the other predictive methods. The final prediction for all plastocyanins by all methods again was chosen to maximize agreement. (See table 1).

The list of ϕ, ψ angles for plastocyanin was also used to construct a tentative three-dimensional model for the molecule. Nicholson molecular models supplied by Labquip, 18 Rosehill Park Estate, Caversham, Reading, England, were used in the construction. Final selection of angle pairs at each residue was based upon the results of the secondary structural predictions, as well as other information. (See below.)

3. Results and discussion

3.1. Secondary structure

Table 1 presents the predictions for the secondary structure of the molecule. There is considerable agreement among all the methods. The final best prediction of secondary structure for plastocyanin, based on information from all four methods, is given in the last column of the table. For predicted β sheet segments shorter than five residues, or α helix residues shorter than six, it is assumed that the residues immediately before and after the predicted segment can participate in the designated structure except in the method of Fasman. For β turns, only the method of Fasman and the set of ϕ, ψ angles (table 2) are applicable. Venkatachalam [15] has specified three classes of turns. Although four residues make up the turn, hydrogen bonding can occur if $C_{\alpha 2}$ and $C_{\alpha 3}$ have one of the following three sets of angle pairs: Class I, −60, −30; −90, 0;; Class II, −60, +120; +80, 0; and Class III, −60, −30; −60, −30. The range must not exceed ±15 degrees for three of the angles, while a fourth angle may deviate by 30 degrees. Since so many residues are in the α helix domain, slight deviations from the α helix values ($\phi = -57$, $\psi = -55$) could easily allow large numbers of Class III (3_{10} helix) turns. Therefore, using an arbitrary cutoff for that class, only angle pairs with $[\phi] > -50$ and $[\psi] < -45$ were chosen. All sets of angles which could possibly give rise to a β turn are indicated in table 2, residues $C_{\alpha 2}$ and $C_{\alpha 3}$ being marked. The number of turns possible is rather large, but they do include all the ones predicted by the Fasman procedure.

3.2. Tertiary structure

If ϕ, ψ angles are available for all residues of a protein, this is sufficient information to construct a three-dimensional model. However, the ϕ, ψ angles obtained here are only statistical preferences, and there are two or three choices for each position. Wu and Kabat [4] tried to predict the torsion angles for cytochrome *c*, using the same data set derived from X-ray analysis (the angles for cytochrome *c* were subtracted out in their study). Out of 101 residues, they obtained good results (±15 degrees maximum deviation) for 59. For another 29 residues, one of the angles deviated by as much as 70 degrees from the X-ray results, and for 17 residues the deviations were more than 100 degrees. Of the 17 poor predictions, 11 were heme contact residues and 6 were for tripeptide combinations not represented in the X-ray data set. In the case of plastocyanin, one might expect slightly better results since the copper is not as large as the heme, and thus should not perturb the folding of the polypeptide chain as much. Further, if the final prediction for secondary structure is taken as correct, there is a unique or nearly unique angle choice (±15° for each angle) for those sequence segments assigned a secondary structure. Finally, for the remainder of the sequence one can arbitrarily choose those angles which occur with the highest frequency in the eleven protein data set. When these assignments are made, eleven tripeptides still do not have a unique or nearly unique pair of angles: 21–22–23, 24–25–26, 32–33–34, 34–35–36, 37–38–39, 46–47–48, 50–51–52, 67–68–69, 87–88–89, 91–92–93, and 92–93–94 (See table 2).

At this point it was decided to construct a three-dimensional representation using Nicholson models and to try both sets of angles for these eleven positions. The final best structure was to be chosen so as to be consistent with other structural properties of the plastocyanin molecule. The relevant properties are: (a) Plastocyanin is a soluble, globular molecule [16]. (b) The copper is internal and is probably coordinated to CYS 85 and up to four other protein residue ligands [16,17] two of which may be HIS [18]. (c) At least one of the three tyrosines is in a medium of low polarity [16]. (d) Since the two conserved regions of the polypeptide chain are 32–48 and 83–95 [7], one would expect that these two regions are involved intimately in copper coordination and other functional aspects. Since it has also been shown that protein evolution occurs mainly at protein surface and external sites [19], one would expect that these two conserved regions would be internal in the molecule. Lastly, (e) from general properties of globular proteins, one would expect that hydrophobic residues would be mostly internal, while hydrophilic residues would be surface or external. Also, some of the β sheets would be expected to form hydrogen bonds with each other.

During the construction it was discovered that the first choice angles occasionally caused the polypeptide chain to form local close contacts. In these cases alterna-

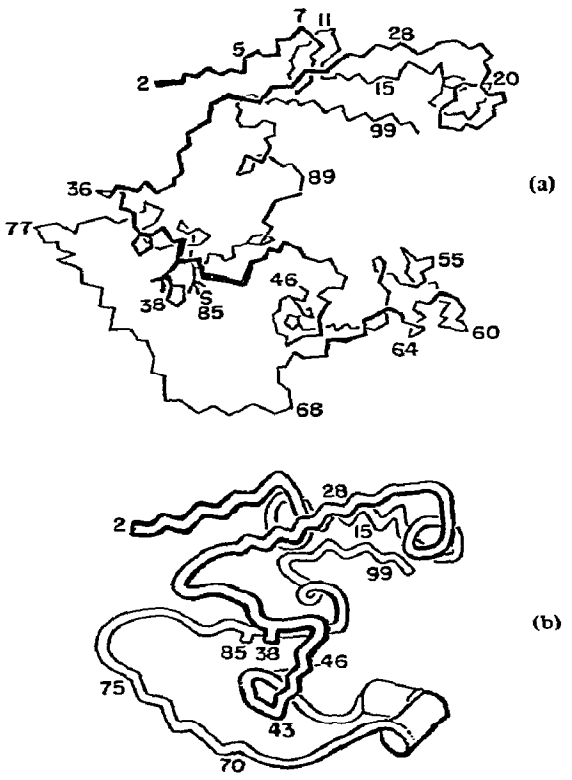

Fig. 1. A possible preliminary tertiary structure for plastocyanin.

Fig. 1a depicts the α-carbon skeleton of the best model obtainable using the ϕ, ψ angles predicted by the method of Wu and Kabat [4]. For the model, the sequence of potato plastocyanin was used, with ϕ, ψ angles identified in table 2 by asterisk. All prolines are shown, as well as the side chains of CYS 85 and HIS 38. The figure was drawn from a three-dimensional simulation constructed with Nicholson models.

Fig. 1b is a schematic drawing of fig. 1a, showing the placement of the predicted secondary structures. Cylinders represent helices, and zig-zag segments represent β sheets. Both models are slightly expanded (less compact) to facilitate representation of chain folding.

tive angle choices were required to obtain a chain free of sterically prohibited interactions. Such an example is tripeptide 33–34–35, in which the angles −88, +154 had to be used instead of the "first choice", −67, −29. The choices used for the final model are indicated in table 2 by an asterisk. The "best" model itself is depicted in fig. 1. Not surprisingly, of the eleven alternative tripeptide angles to be tested, those in the center of the polypeptide chain, such as at 46–47–48 and at 67–68–69, had the greatest effect. This model conforms to most of the criteria listed above: Conserved residues 34–47 and 81–98 are internal. The copper site is internal, and the major coordinating ligands are CYS 85, ASP 45, GLU 46, and HIS 38; involvement with LYS 78, THR 74 and SER 72 is also a possibility. HIS 88 is rigidly fixed too far away from CYS 85 (9 Å between imidazole N and sulfhydryl S) to participate. All the angle choices for residues 85–88 were tested in an attempt to obtain a closer approach. Within the framework of the other restrictions, especially the internal positions of 32–48 and 83–95, no closer approach was possible. (Residues 38 and 88 are the two invariant HIS in plastocyanin.) TYR 81 and TYR 84 are in the copper environment, while TYR 71 is at a surface location. In general, hydrophilic residues are accessible to solvent (the only exception appears to be ASP 62), and hydrophobic residues are internal or can rotate to the internal side of the polypeptide chain. Four of the six β sheets are able to form hydrogen bonds with each other. Sheet 2–6 appears to be able to form three bonds with sheet 27–32 (LEU 5 amide – PHE 30 carbonyl, LEU 5 carbonyl – PHE 30 amide, and GLY 7 amide – ILE 28 carbonyl). Also, β sheet 14–17 forms four hydrogen bonds with sheet 96–99 (ALA 14 carbonyl – VAL 97 amide, ALA 14 amide – VAL 97 carbonyl, PHE 15 carbonyl – VAL 99 amide, and GLY 18 amide – VAL 99 carbonyl). The points at which the main-chain bends (not necessarily β bends) occur are 15–18, 36–39, 47–50, 59–62, 67–70, 76–79, 87–90, and 93–96. The $C_{\alpha 1}$–$C_{\alpha 4}$ distance in all these bends is less than 5.7 Å [20], as well as in 7–10. Hydrogen bonding between $C_{\alpha 1}$ and $C_{\alpha 4}$ is apparently possible in all but 36–39. Lim's designation of internal and surface β sheets [2] does not correspond to the results of this model (See table 1). Whether the general disposition of chain folding in this model resembles that of plastocyanin must await the results of X-ray analysis.

A more thorough treatment of the problem would require interactive computer graphics (since a Nicholson model of this size is quite awkward to manipulate) and energy minimization techniques. Further, ϕ, ψ angle data is now available for several proteins not included in Wu and Kabat's original set of eleven, and the computer search for torsion angle preferences

should be repeated with this larger data set.

Acknowledgements

The author is deeply indebted to Dr. A.T. Linde of the Carnegie Institution of Washington for writing the computer program for the ϕ, ψ angle search. Thanks also to Professor D. Boulter for permitting the examination of the amino acid sequences of shepherd's purse, marrow, and spinach plastocyanin prior to publication; to Professor R.C. Lewontin for providing space and facilities to complete this work; and to Drs. E.A. Kabat and D. Wiley for helpful advice and discussion.

References

[1] P. Chou and G. Fasman, Biochemistry 13 (1974) 222.
[2] V.I. Lim, J. Mol. Biol. 88 (1974) 873.
[3] E.A. Kabat and T.T. Wu, Biopolymers 12 (1973) 751.
[4] T.T. Wu and E.A. Kabat, J. Mol. Biol. 75 (1973) 13.
[5] J. Kelly and R.P. Ambler, Biochem. J. 143 (1974) 681.
[6] P.R. Milne, J.R.E. Wells and R.P. Ambler, Biochem. J. 143 (1974) 690.
[7] J.A.M. Ramshaw, M.D. Scawen and D. Boulter, Biochem. J. 141 (1974) 835.
[8] M.D. Scawen and D. Boulter, Biochem. J. 143 (1974) 257.
[9] M.D. Scawen, J.A.M. Ramshaw and D. Boulter, Biochem. J. 147 (1975) 343.
[10] M.D. Scawen and D. Boulter, unpublished experiments.
[11] J. Kelly, M. Sc. Thesis, University of Edinburgh (1971).
[12] J.A.M. Ramshaw, M.D. Scawen, C.J. Bailey and D. Boulter, Biochem. J. 139 (1974) 583.
[13] R.P. Ambler, in: Recent Developments in the Chemical Study of Protein Structure, eds. A. Previero, J.-F. Pechere, and M.-A. Coletti-Previero (Inserm, Paris, 1971) p. 289.
[14] B. Honig, E.A. Kabat, L. Katz, C. Levinthal and T.T. Wu, J. Mol. Biol. 80 (1973) 277.
[15] C.M. Venkatachalam, Biopolymers 6 (1968) 1425.
[16] M.T. Graziani, A. Finazzi-Agro, G. Rotilio, D. Barra and B. Mondovi, Biochemistry 13 (1974) 804.
[17] V. Miskowski, S.-P. W. Tang, T.G. Spiro, E. Shapiro and T.H. Moss, Biochemistry 14 (1975) 1244.
[18] J.L. Markley, E.L. Ulrich, S.P. Berg and D.W. Krogmann, Biochemistry 14 (1975) 4428.
[19] A.B. Champion, K.L. Soderberg and A.C. Wilson, J. Molec. Evolution 5 (1975) 291.
[20] J.L. Crawford, W.N. Lipscomb and C.G. Schellman, Proc. Nat. Acad. Sci. U.S. 70 (1973) 538.