Journal of Protein Chemistry, Vol. 15, No. 2, 1996

Prediction of Protein Secondary Structure from Amino Acid Sequence

Jen Tsi Yang[1]

Received November 6, 1995

The conformational parameters P_k for each amino acid species ($j = 1$–20) of sequential peptides in proteins are presented as the product of $P_{i,k}$, where i is the number of the sequential residues in the kth conformational state ($k = \alpha$-helix, β-sheet, β-turn, or unordered structure). Since the average parameter for an n-residue segment is related to the average probability of finding the segment in the kth state, it becomes a geometric mean of $(P_k)_{av} = \prod (P_{i,k})^{1/n}$ with amino acid residue i increasing from 1 to n. We then used $\ln(P_k)_{av}$ to convert a multiplicative process to a summation, i.e., $\ln(P_k)_{av} = (1/n) \sum P_{i,k}$ ($i = 1$ to n) for ease of operation. However, this is unlike the popular Chou–Fasman algorithm, which has the flaw of using the arithmetic mean for relative probabilities. The Chou–Fasman algorithm happens to be close to our calculations in many cases mainly because the difference between their P_k and our $\ln P_k$ is nearly constant for about one-half of the 20 amino acids. When stronger conformation formers and breakers exist, the difference become larger and the prediction at the N- and C-terminal α-helix or β-sheet could differ. If the average conformational parameters of the overlapping segments of any two states are too close for a unique solution, our calculations could lead to a different prediction.

KEY WORDS: Chou–Fasman algorithm; protein primary structure; secondary structure prediction.

1. INTRODUCTION

The amino acid sequence of a native protein carries molecular information that determines the three-dimensional structure of the protein (secondary and tertiary structures). At present protein folding has been actively studied both theoretically and experimentally; it is one of the important subjects in protein chemistry. However, because of the complexity of protein molecules, we are still a long way from understanding and predicting protein structure with certainty.

During the past two decades more than 20 empirical methods have been proposed for predicting protein secondary structure (α-helix, β-sheet, β-turn, and unordered structure) (Fasman, 1989*a*, *b*). Among these methods the approaches of Chou and Fasman (1974*a*, *b*) and of Garnier *et al.* (1978) and Garnier and Robson (1989) and the stereochemical method of Lim (1974*a*, *b*) have been frequently used. The very simplicity of the Chou–Fasman method has made this algorithm easy to understand and use. It has been described by Mathews and van Holde (1990) in their popular textbook on biochemistry. Its statistical calculations of the conformational parameters for 20 amino acid species are straightforward. Nevertheless, the averaging of these parameters by the arithmetic mean is a serious flaw because a combination of relative probabilities should be multiplicative, not additive. This problem was overlooked by the authors, and it has remained for 20 years now. Perhaps this is a case of "end justifies means," because the Chou–Fasman prediction has so far probably had an overall accuracy of up to 70%. Nevertheless doubts still exist about the above-

[1] Cardiovascular Research Institute, University of California, San Francisco, California 94143-0130.

0277-8033/96/0200-0185$09.50/0

mentioned flaw. I will attempt to explain this puzzle: under certain conditions this flaw does not happen to affect the prediction too much, but under other conditions it could lead to different predictions for protein secondary structure. By taking logarithms of conformational parameters, the operations are converted from the geometric mean to the arithmetic mean for the ease of calculation, but are still on a secure foundation.

2. THE CHOU–FASMAN ALGORITHM

The relative frequencies of 20 amino acid species in α-helix, β-sheet, β-turns, and unordered structure of a set of globular proteins are calculated from their occurrences in three-dimensional structure based on X-ray crystallographic data. Chou and Fasman (1974*a*, *b*; 1977) first used 15 globular proteins with 2473 amino acid residues, which were then expanded to 29 proteins with 4741 residues (Chou and Fasman, 1978*a*, *b*), and, more recently, further updated to 64 proteins with 11 445 residues (Chou, 1979, 1989). The latter included 19α, 14β, 14$\alpha + \beta$, and 16α/β proteins, thus better representing the four classes of proteins than the previous sets of proteins. However, β-turns were still counted from the previous 29 proteins, which already included 459 β-turns.

The frequency of occurrence of amino acid species j in the kth conformational state of proteins is

$$f_{j,k} = n_{j,k}/n_j \tag{1}$$

where n refers to the number of residues. Since a segment of polypeptide chain in a protein often has the potential of forming more than one conformation state, Chou and Fasman introduced the conformational parameter $P_{j,k}$ for amino acid species j in the kth conformational state, which is defined as

$$P_{j,k} = f_{j,k}/\langle f_k \rangle \tag{2}$$

and

$$\langle f_k \rangle = \sum_{j=1}^{20} n_{j,k} \Big/ \sum_{j=1}^{20} n_j \tag{3}$$

Here Eq. (3) represents the total number of amino acid residues in the kth conformational state for the chosen set of proteins. It is a constant and depends to some extent on the set of chosen proteins. It was 34.8% α-helix and 24.9% β-sheet for 64 proteins studied (Chou, 1989) and 38.7% β-turn for 29 proteins studied (Chou and Fasman, 1977). To calculate $P_{j,k}$ in Eq. (2), $f_{j,k}$ in Eq. (1) is normalized by dividing it by $\langle f_\alpha \rangle$, $\langle f_\beta \rangle$, and $\langle f_t \rangle$, respectively, which are, respectively, 0.348, 0.249, and 0.387. This makes it possible to compare the three conformational states for a segment of sequential peptides and choose one state as the predicted state by following certain rules. Thus, $P_{j,k}$ is the frequency of species j in the kth state divided by a coefficient; its numerical value can be greater or less than one for a strong conformation former or breaker, respectively (although the frequencies or probabilities are always less than or equal to one).

To avoid duplication, the rules for the Chou–Fasman algorithm will be included in the next section. In their algorithm Chou and Fasman chose the arithmetic mean for the conformational parameters $\langle P_k \rangle$ for a segment of amino acid sequence, that is,

$$\langle P_k \rangle = (1/n) \sum_{i=1}^{n} P_{i,k} \tag{4}$$

3. THE IMPORTANCE OF GEOMETRIC MEAN

The average probability of finding a segment of sequential peptide in the kth conformational state is a geometric mean; thus, for n residues of a segment in the kth state, the average frequency of occurrence is the nth root of the produce of $f_{i,k}$, or

$$[(f_k)_{\text{av}}]^n = \prod_{i=1}^{n} f_{i,k} \tag{5}$$

in contrast to the summation in Eq. (4). [To avoid confusion with using the same symbol $\langle \cdot \rangle$ as in Chou and Fasman (1974*a*), we will use the symbol $(\cdot)_{\text{av}}$ to represent the average of a product of multiple probabilities.] Each ith residue belongs to one of the 20 amino acid species (j). To follow the chou–Fasman algorithm, each $f_{j,k}$ can be divided by a constant $(f_{j,k})_{\text{av}}$ and converted to $P_{j,k}$ [see Eq. (2)]. Thus, we have

$$[(P_k)_{\text{av}}]^n = \prod_{i=1}^{n} [f_{i,k}/(f_k)_{\text{av}}] \tag{6}$$

and

$$\ln(P_k)_{\text{av}} = (1/\text{n}) \sum_{i=1}^{\text{n}} \ln P_{i,k} \tag{7}$$

after taking the logarithm of both sides of Eq. (5). This converts Eq. (6) from multiplication to addition. While Eq. (4) resembles Eq. (7) by taking an arithmetic mean, $\ln(P_k)_{\text{av}}$ is based on a solid foundation for the combination of probabilities.

The Chou–Fasman prediction happens to coincide with our calculation only when the difference between P_k in Eq. (4) and $\ln P_k$ in Eq. (7) is the same for all species j. Under these circumstances only the cutoff values will be altered from P_k to $\ln P_k$. For instance, if $(P_k - \ln P_k) = 1$ for each amino acid residue of a segment, then $\langle P_k \rangle - \ln(P_k)_{av}$ would be unity, too. But the cutoff will be changed from $\langle P_k \rangle = 1$ in the Chou–Fasman algorithm to $\ln(P_k)_{av} = 0$.

Table I lists the $100(P_k - \ln P_k)$ values for α-helix, β-sheet, and β-turn of the 20 amino acid species in proteins. In all three cases the numerical values of five or six rows in the middle section are 100 (i.e., $P_k - \ln P_k = 1.00$); that is, the difference remains unchanged. The differences become larger for rows above and below this middle section. The strongest conformation breakers for α-helix, β-sheet, and β-turn are 115, 118, and over 120, respectively, whereas the strongest conformation formers are under 110 for α-helix and around 110 for the other two states. These findings can be seen more clearly and explained by expanding the term $\ln P_k$ into a series of $(P_k - 1)$ so that we have

$$P_k - \ln P_k = 1 + (P_k - 1)^2/2 - (P_k - 1)^3/3 + \cdots, 2 > P_k > 0 \quad (8)$$

Table I. Differences in Conformational Parameters of Amino Acid Species in Globular Proteins[a]

$100(P_\alpha - \ln P_\alpha)$		$100(P_\beta - \ln P_\beta)$		$100(P_t - \ln P_t)$	
Glu	107	Val	114	Asn	111
Ala	106	Ile	112	Gly	111
Met	104	Thr	104	Pro	110
Leu	104	Tyr	104	Asp	108
Lys	102	Trp	102	Ser	107
His	101	Phe	102	Cys	102
Gln	101	Leu	101	Tyr	101
Phe	101	Cys	100	Lys	100
Asp	100	Met	100	Gln	100
Trp	100	Gln	100	Thr	100
Arg	100	Ser	100	Trp	100
Ile	100	Arg	100	Arg	100
Val	100	Gly	101	Arg	100
Cys	100	HIS	102	His	100
Thr	103	Ala	103	Glu	104
Asn	103	Lys	105	Ala	108
Tyr	105	Asp	108	Met	111
Ser	105	Asn	108	Phe	111
Gly	109	Pro	110	Leu	112
Pro	115	Glu	118	Val	119
				Ile	123

[a] Calculated from the updated P_k values by Chou (1979); see text for details.

Evidently, the term $100(P_k - \ln P_k)$ becomes 100 when $P_k = 1$ (cf. Table I). If P_k is not too far from unity, this term will still be close to 100. On the other hand, stronger conformation formers and, in particular, stronger conformation breakers will widen the difference between P_k and $\ln P_k$, more so toward the top and bottom rows in Table I. In all cases P_k is still positive and less than 2. As long as these differences between P_k and $\ln P_k$ for the 20 amino acids do not vary too much, the Chou and Fasman arithmetic mean of conformational parameters should fortuitously give similar predictions to those based on the arithmetic mean of $\ln P_\alpha$, $\ln P_\beta$, and $\ln P_t$. But the data in Table I indicate that initiation and, in particular, termination of the propagation of α-helices and β-sheets might be affected to some extent according to our modification.

4. THE RULES OF PREDICTION

We used the same Chou–Fasman rules for predicting protein secondary structure, except that P_α, P_β, P_t, and p_t are replaced by their logarithmic terms (Table II). They can be briefly summarized as follows [for details of the prediction rules, see Chou and Fasman (1974*a*, *b*)]:

Table II. Conformational Parameters of Amino Acid Species for α-Helix, β-Sheet, and β-Turn of Globular Proteins

Residue	$100 \ln P_\alpha$	$100 \ln P_\beta$	$100 \ln P_t$
Ala	32.9	−23.6	−41.6
Arg	0	−6.2	−5.1
Asn	−24.8	−41.6	44.5
Asp	5.8	−41.6	37.8
Cys	−5.1	6.8	17.4
Gln	11.3	0	−2.0
Glu	36.5	−67.3	−30.1
Gly	−46.2	−13.9	44.5
His	11.3	−18.6	−5.1
Ile	−1.0	45.1	−75.5
Leu	26.2	15.7	−52.8
Lys	19.1	−31.5	1.0
Met	27.8	1.0	−51.1
Phe	10.4	20.7	−51.1
Pro	−59.8	−47.8	41.9
Ser	−32.9	−6.2	35.8
Thr	−24.8	28.5	−4.1
Trp	3.0	21.5	−4.1
Tyr	−31.5	27.0	13.1
Val	−3.0	49.5	−69.3

1. Four α-helix formers out of six amino acid residues will nucleate an α-helix, which is extended in both directions until α-tetrapeptide breakers with $\ln(P_\alpha)_{av} < 0$ are reached [see Eq. (7)]. The segment is predicted to be α-helix if log(arithmetic mean of P_α) or $\ln(P_\alpha)_{av} > 0.03$. Further, for segments having both α-helix and β-sheet potentials, $\ln(P_\alpha)_{av} > \ln(P_\beta)_{av}$ for the overlapping portion. [The Chou and Fasman cutoff is at $\langle P_\alpha \rangle = 1.03$, so that $\ln(P_\alpha)_{av} = 0.03$.]

2. Three β-sheet formers out of five residues will initiate a β-sheet, which is extended in both directions until β-tetrapeptide breakers with $\ln(P_\beta)_{av} < 0$ are reached. The segment is predicted to be β-sheet if log(arithmetic mean of P_β) or $\ln(P_\beta)_{av} > 0.05$ as well as $\ln(P_\beta) > \ln(P_\alpha)$ for overlapping portion, if any. [The Chou and Fasman cutoff is at $\langle P_\beta \rangle = 1.05$, so that $\ln(P_\beta)_{av} = 0.05$.]

3. Any tetrapeptide segment is predicted to be β-turn if $\ln(p_t)_{av} > -2.4$ as well as $\ln(P_\alpha)_{av} < \ln(P_t)_{av} > \ln(P_\beta)_{av}$ for the overlapping portions. Chou and Fasman (1977) defined p_t as $f_1 f_2 f_3 f_4$ [Table III; see also Eq. (1)] for the probabilities of finding the residues at the first, second, third, and fourth positions, and $p_t > 0.75 \times 10^{-4}$. We will again use the geometric mean by defining $[(p_t)_{av}]^4 > 0.75 \times 10^{-4}$ or $\ln(p_t) > -2{\cdot}4$, so that $\ln(p_t)_{av}$ represents the average p_t of the probabilities at the four positions in a manner similar to that of Eq. (7). That is, by taking the logarithms of both sides, we have

$$\ln(p_t) = \sum_{i=1}^{4} \ln f_i \qquad (9)$$

again converting the product of probabilities to a summation. Only additions are required throughout the calculations as $\ln(P_\alpha)_{av}$, $\ln(P_\beta)_{av}$, and $\ln(P_t)_{av}$ are.

The lowercase p_t for the probabilities should not be confused with the capital P_t for the normalized conformational parameter. Chou and Fasman (1974*a*, *b*) did multiply the four frequencies of occurrence f_i for the four residues in a β-turn, but then reverted to the use of arithmetic means for calculating the average conformational parameters for the β-turn as they did for α-helix and β-sheet. In our calculations we have also raised p_t to the fourth power. Throughout our treatment we have consistently used the geometric mean for the calculation of all conformational parameters and also the average probability of four residues in a β-turn.

Table III. β-Turn Residues in the Four Positions Based on 29 Proteins[a]

$\ln f_1$		$\ln f_2$		$\ln f_3$		$\ln f_4$	
Asn	−1.8	Pro	−1.2	Asn	−1.7	Trp	−1.8
Cys	−1.9	Ser	−2.0	Gly	−1.7	Gly	−1.9
Asp	−1.9	Lys	−2.2	Asp	−1.7	Cys	−2.1
His	−2.0	Asp	−2.2	Ser	−2.1	Tyr	−2.1
Ser	−2.1	Thr	−2.2	Cys	−2.2	Ser	−2.2
Pro	−2.3	Arg	−2.2	Tyr	−2.2	Gln	−2.3
Gly	−2.3	Gln	−2.3	Arg	−2.3	Lys	−2.4
Thr	−2.5	Gly	−2.5	His	−2.4	Asn	−2.4
Tyr	−2.5	Asn	−2.5	Glu	−2.6	Arg	−2.5
Trp	−2.6	Met	−2.5	Lys	−2.6	Asp	−2.5
Gln	−2.6	Ala	−2.6	Thr	−2.7	Thr	−2.5
Arg	−2.7	Tyr	−2.7	Phe	−2.7	Leu	−2.7
Met	−2.7	Glu	−2.8	Trp	−2.8	Pro	−2.7
Val	−2.8	Cys	−2.9	Gln	−3.3	Phe	−2.7
Leu	−2.8	Val	−3.0	Leu	−3.3	Glu	−2.8
Ala	−2.8	His	−3.1	Ala	−3.4	Ala	−2.9
Phe	−2.8	Phe	−3.2	Pro	−3.4	Ile	−2.9
Glu	−2.9	Ile	−3.4	Val	−3.6	Met	−2.9
Lys	−2.9	Leu	−3.7	Met	−4.3	His	−2.9
Ile	−3.2	Trp	−4.3	Ile	−4.3	Val	−2.9

[a] Calculated from the f_i values in Chou and Fasman (1977).

5. TEST OF THE MODIFIED METHOD

As an example we will compare the structural assignment of the 150-residue staphylococcal nuclease by the Chou–Fasman method (Prevelige and Fasman, 1989) and that based on our use of $\ln(P_k)_{av}$. Table IV lists the first 20 amino acid residues for illustration. The data base used by Prevelige and Fasman (1989) was 29 proteins, but their output file was essentially the same as that with 64 proteins shown in Table IV. Further, the predictions based on 29 and 64 proteins were virtually identical with each other for all 159 amino acid residues (data not shown). In the Chou and Fasman method the cutoffs for their $\langle P_\alpha \rangle$, $\langle P_\beta \rangle$, and $\langle P_t \rangle$ of tetrapeptides were lowered to 1.00. We have therefore defined the asterisks after an amino acid residue as the one whose $\ln(P_k)_{av}$ for its sequential tetrapeptide was equal to or greater than zero (i.e., $\ln 1 = 0$).

Frequently a segment displays a propensity for more than one kth state and is hard to call, and a potential α-helix or a β-sheet may also be linked to a potential β-turn. To use the Chou–Fasman protocol one first looks for β-turns based on the asterisks after $\ln(p_t)_{av}$ (see Table IV) and then

Table IV. The Algorithm for the First 20 Amino Acid Residues of Staphylococcal Nuclease[a]

Residue	$100\langle P_\alpha\rangle$	$100\ln(P_\alpha)_{av}$	$100\langle P_\beta\rangle$	$100\ln(P_\beta)_{av}$	$100\langle P_t\rangle$	$100\ln(P_t)_{av}$	$\ln(f_t)_{av}$
1 Ala	91	−12	109*	7*	100	−3.5	−2.4
2 Thr	87	−16	108*	5*	109	7.2	−2.4
3 Ser	98	−5	93	−9	110	8.4	−2.3*
4 Thr	112*	9*	99	−4	89	−13.5	−2.5
5 Lys	121*	18*	86	−16	89	−14.0	−2.8
6 Lys	121*	18*	86	−16	89	−14.0	−2.8
7 Leu	126*	23*	81	−25	82	−21.8	−2.8
8 His	108*	1*	67	−41	105	1.8	−2.4*
9 Lys	114*	7*	66	−42	98	−7.2	−3.0
10 Glu	104*	−4	81	−27	97	−8.5	−2.5
11 Pro	100*	−6	97	−6	93	−14.2	−2.6
12 Ala	111*	8*	121*	16*	67	−43.5	−2.8
13 Thr	107*	4*	120*	14*	75	−32.9	−3.2
14 Leu	122*	19*	106*	1*	68	−42.2	−2.9
15 Ile	114*	12*	116*	8*	65	−47.9	−2.9
16 Lys	116*	14*	93	−13	90	−19.6	−3.1
17 Ala	102*	−1	97	−8	103	−8.7	−2.5
18 Ile	93	−8	94	−13	123	11.2	−9.5
19 Asp	88	−14	88	−17	136	29.0	−2.2*
20 Gly	86	−16	112*	5*	112	2.2	−2.6

[a] The $\langle P_k\rangle$ values were taken from Prevelige and Fasman (1989).

compares $\ln(P_t)_{av}$ with $\ln(P_\alpha)_{av}$ and $\ln(P_\beta)_{av}$ for this tetrapeptide (rule 3) (Fasman, G. D., personal communication). When a series of β-turns is found to overlap, the turn with the higher local $\ln(p_t)_{av}$ is assigned as β-turn. However, one should be certain that no potential β-turns have been discounted. If one or two residues of the turn are also linked to an α-helix or a β-sheet, such residue or residues are considered as part of the β-turn only.

In the example, Prevelige and Fasman (1989) started by assigning residues 3–6 of staphylococcal nuclease as β-turn because their $p_t > 0.75 \times 10^{-4}$ (not shown) as well as $\langle P_\alpha\rangle < \langle P_t\rangle > \langle P_\beta\rangle$ (see Table IV). Residues 7–18 were assigned as α-helix, as their $\langle P_\alpha\rangle$ equaled 1.13. Residues 8–11 were not a β-turn, since $\langle P_t\rangle$ was smaller than $\langle P_\alpha\rangle$. Pro-11 cannot be located inside the inner or C-terminal helix, and therefore it may form a bend in the long helical segment. For residues 12–15, $\langle P_\beta\rangle = 1.21$ and $\langle P_\alpha\rangle = 1.11$, but this tetrapeptide was considered to be too short for a β-sheet when compared with the overlapping α-helix, which could be extended in both directions; thus, Prevelige and Fasman (1989) assigned it as part of the 12-residue α-helix segment.

We followed the Chou and Fasman procedure by looking for the asterisks for β-turns first. Residues 3–6 gave $\ln(p_t)_{av} = -2.3$ (which is less negative than the cutoff of −2.4) and thus it had the potential of forming a β-turn. Further, $\ln(P_t)_{av}$ was also in accord with this assignment because it was greater than either $\ln(P_\alpha)_{av}$ or $\ln\langle P_\beta\rangle$. Our assignment seems to have a few possible differences from that of Prevelige and Fasman (1989). Residues 8–11 of the nuclease could be another β-turn (Table IV). For this tetrapeptide $\ln(p_t)_{av}$ was −2.4, on the borderline of the cutoff, but $\ln(P_\alpha)_{av} < \ln(P_t)_{av} > \ln(P_\beta)_{av}$. From Table IV we can see that residues 8–11 had $\ln(P_\alpha)_{av} = 1.0$, which was less than $\ln(P_t)_{av} = 1.8$, whereas the corresponding Chou and Fasman $\langle P_\alpha\rangle = 1.08$ was greater than $\langle P_t\rangle = 1.05$. Further, $\ln(P_\alpha)_{av}$ for residues 10–13 was a tetrapeptide breaker. These findings led to a different assignment from the Chou–Fasman algorithm, which included residues 8–11 in the α-helix segment. Thus, our assignment can differ from that of Chou and Fasman when any average conformational parameters for the overlapping segments are close to each other, as in the present case for residues 8–11. Residues 12–18 could be assigned as α-helix, which was linked to a second β-turn for residues 19–22 (not shown), just as Prevelige and Fasman (1989) did. Actually, excluding residues 8–11 from the α-helix segment better fits the prediction with X-ray crystallographic data, which indicate a helical segment for residues 12–19, but not for residues 7–11 (Prevelige and Fasman, 1989).

Thus, *the Chou–Fasman method does not provide a unique solution to any segment of a polypeptide chain, and the assignments are often up to the discretion of the user.* However, Prevelige and Fasman (1989) pointed out that significant information will be lost if the algorithm is reduced to a single predicted structure. They further stipulated that the "hard to call" region may be sites of conformational changes. "By being forced to be in closer contact with the raw data, the user will become aware of these potentialities." I believe that the present algorithm can provide a reasonable prediction of protein secondary structure and its results should complement those of other physical techniques such as circular dichroism of proteins. The method can also be extended to the study of modification and denaturation of a native protein once the latter's secondary structure can be predicted with confidence. But as with any empirical method, our calculations should still be interpreted with caution.

6. CONCLUDING REMARKS

The main goal of this work is to correct a logical flaw in the Chou–Fasman method of predicting protein secondary structure. Since the probabilities of finding amino acid residues of a segment in the kth conformational state ($k = \alpha$-helix, β-sheet, β-turn, and unordered structure) can only be combined through multiplication, we use the geometric mean to express the averaging of conformational parameters. For ease of calculations we take logarithms of the average P_k of n residues in the segment and convert a geometric mean to an arithmetic mean, that is, $\ln(P_k)_{av} = (1/n)\sum \ln P_k$. To be consistent with our definition, the average probability of finding a β-turn with four residues is also expressed as $\ln(p_t)_{av} = (\sum \ln f_i)/4$. (Because the conformational parameter $P_{j,k}$ is obtained by dividing the frequency of occurrence of amino acid species j by a constant for the average frequency of occurrence of all residues in the kth state, the numerical values of $P_{j,k}$ can be greater than one, although the probabilities per se always are less than or equal to one.)

Biochemistry is sometimes considered to be "old-fashioned," but it still holds everything in molecular biology together. This work emphasizes the "old-fashioned" physical principle that averaging relative probabilities is multiplicative, not additive. Our purpose is not to improve the accuracy of the Chou–Fasman algorithm by a few percent. In fact, our calculations could lead to a less desirable prediction, but they are based on a more solid foundation. On the other hand, if the Chou–Fasman prediction happens to agree with our calculations, we must seek an explanation, as is illustrated in this work. However, the logarithms of average conformational parameters (Tables II and III) should be used for the prediction.

A protein molecule contains segments of α-helix, β-sheet, and β-turn. The Chou–Fasman method often overestimates or underestimates individual segments. However, even if its accuracy drops from 70% to 50%, this does not mean that the prediction resembles the flip of a coin. The ease of calculations within a matter of hours makes the method very attractive. A prediction is just a prediction, be it the Chou–Fasman algorithm or any other recent prediction from protein folding studies. (The three-dimensional structure of a protein is now solved by X-ray diffraction and NMR methods.) The modified Chou–Fasman method as described in this work can continue to be used until a more accurate and routine method becomes available.

ACKNOWLEDGMENTS

I thank Prof. L. Peller for his critical comments on the manuscript and a stimulating discussion, and Qing-Min Wang for her assistance in running the computer program.

REFERENCES

Chou, P. Y. (1979). New approaches to protein structural analysis and conformational prediction, in *Rapport d'Activité Scientifique du CECAM*, Orsay, pp. 149–216.

Chou, P. Y. (1989). Prediction of protein structural classes from amino acid compositions, in *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G. D., ed.), Plenum Press, New York, pp. 549–586.

Chou, P. Y., and Fasman, G. D. (1974*a*). Conformational parameters for amino acids in helical, β-$\sigma\eta\varepsilon\varepsilon\tau$, and random coil regions calculated from proteins, *Biochemistry* **13,** 211–222.

Chou, P. Y., and Fasman, G. D., (1974*b*). Prediction of protein conformation, *Biochemistry* **13,** 222–245.

Chou, P. Y., and Fasman, G. D. (1977). β-turns in proteins, *J. Mol. Biol.* **115,** 135–175.

Chou, P. Y., and Fasman, G. D. (1978*a*). Prediction of the secondary structure of proteins from their amino acid sequence, *Adv. Enzymol.* **47,** 45–148.

Chou, P. Y., and Fasman, G. D. (1978*b*). Empirical prediction of protein conformation, *Annu. Rev. Biochem.* **47,** 251–276.

Fasman, G. D., ed. (1989*a*). *Prediction of Protein Structure and*

the Principles of Protein Conformation, Plenum Press, New York.

Fasman, G. D. (1989*b*). The development of the prediction of protein structure, in *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G. D., ed.), Plenum Press, New York, pp. 193–316.

Garnier, J., and Robson, B. (1989). The GOR method for predicting secondary structures in proteins, in *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G. D., ed.), Plenum Press, New York, pp. 417–465.

Garnier, J., Osguthorpe, D. J., and Robson, B. (1978). Analysis of the accuracy and implication of simple methods for predicting the secondary structure of globular proteins, *J. Mol. Biol.* **120,** 97–120.

Lim, V. I. (1974*a*). Structural principles of the globular organization of protein chains: A stereochemical theory of globular protein secondary structure, *J. Mol. Biol.* **88,** 857–872.

Lim, V. I. (1974*b*). Algorithms for prediction of α-helices and β-structure in globular proteins, *J. Mol. Biol.* **88,** 873–894.

Mathews, C. K., and van Holde, K. E., (1990). *Biochemistry,* Benjamin/Cummings, Redwood City, California.

Prevelige, P., Jr., and Fasman, G. D., (1989). Chou–Fasman prediction of the secondary structure of proteins. The Chou–Fasman–Prevelige algorithm, in *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G. D., ed.), Plenum Press, New York, pp. 391–416.