

蛋白质的二级结构预测

——Chou 和 Fasman 方法

华庆新

(中国科学院,生物物理研究所,北京)

蛋白质分子的一切高级结构,都由一级结构即氨基酸残基序列所包含的信息决定。多年来,由蛋白质的氨基酸序列预测二级结构的方法不下十几种。其中,Chou 和 Fasman 的方法自 1974 年提出^[1,2],至 1978 年修正、精化^[3,4],已得到了很好结果,越益受到重视^[5,6]。此方法的突出优点是简便,无须计算机的复杂分析,就可预测出蛋白质的二级结构,准确性约为 80%。目前蛋白质二级结构的测定,当然以 X-晶体衍射结果最准确。Chou 和 Fasman 方法正是基于晶体分析的结果,经统计得出的一整套数据和预测规则。在溶液态,远紫外圆二色性(CD)可测定蛋白质分子的二级结构。近来,二维核磁共振技术(2D-MNR)在测定二级结构方面也开始崭露头角^[7]。CD 方法虽简便易测,但它只能了解各二级结构成份的相对比例。既不知道有几段螺旋或 β 结构,更不知道每段二级结构从哪个残基开始到哪个残基为止。而 Chou 和 Fasman 的序列结构预测方法可帮助分析圆二色结果,推测出二级结构的某些细节。

Chou 和 Fasman 方法的基本思想是由已知晶体结构的蛋白质分子,统计出各种氨基酸残基在螺旋、 β -结构、 β -回折和无规卷曲态中的几率,然后由各氨基酸残基的相应几率来推算各肽段形成二级结构的可能,再根据某些统计规则确定此肽段应形成什么二级结构。1974 年他们对当时已知晶体结构的十五种蛋白进行统计,即羧肽酶 A, α -胰凝乳蛋白酶,细胞色素 b_5 , 弹性蛋白酶、高铁细胞色素 c, α 和 β -血红蛋白(牛),胰岛素,溶菌酶,肌浆蛋白,肌红蛋白,木

瓜蛋白酶,核糖核酸酶,葡萄糖球菌核酸酶和枯草杆菌蛋白酶^[1]。此后,又加进了四种蛋白,即伴刀豆球蛋白,八目鳗血红蛋白,嗜热菌蛋白酶,胰蛋白酶抑制剂的统计数据^[2],1978 年他们在十九种蛋白的基础上,又加进了十个已知晶体结构的蛋白,即腺苷酸激酶,碳酸酐酶,染色质高铁蛋白,细胞色素 c_2 ,铁氧还蛋白,黄素氧还蛋白,蚊蠓幼虫血红蛋白,乳酸脱氢酶,红色氧还蛋白,glycera 血红蛋白,得到由 29 个蛋白统计结果,将预测精度大大推进了一步。

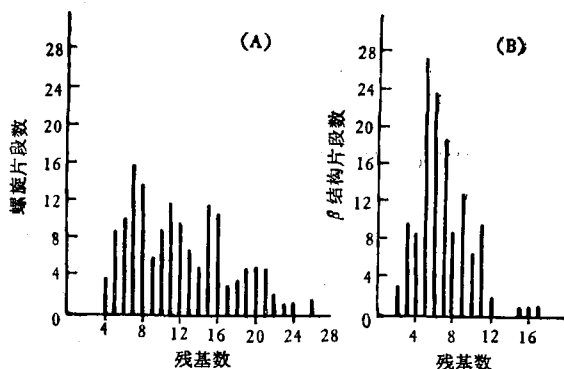


图 1 29 个蛋白中每段螺旋及 β 结构的残基数

表 1 为按 29 个蛋白得出的在螺旋, β 结构、无规卷曲和 β 回折中的残基数目以及所有残基在二级结构中出现的频率。由表 1 可知,在 4741 个残基中,分别有 1798、930 和 1400 个残基分处螺旋(α)、 β -结构(β)和 β -回折(T)中。而 20 种残基在 α , β , 无规卷曲(C)和 β -回折(T)中出现的频率分别为 $\langle f_\alpha \rangle = 0.38$, $\langle f_\beta \rangle = 0.20$, $\langle f_c \rangle = 0.42$, $\langle f_t \rangle = 0.30$ 。其中,每个残基可分处于 α , β , 无规卷曲中,因此 $f_\alpha + f_\beta +$

表 1 29 个蛋白分处 α, β, C, T 区的残基数及
所有残基在二级结构中的平均频率

残基	残基数	α 残基	β 残基	C 残基	T 残基 ^(b) (n_i)
Ala	434	234	71	129	85
Arg	142	53	26	63	40
Asn	230	58	40	132	106
Asp	273	105	29	139	118
Cys	94	25	22	47	33
Gln	162	68	35	59	47
Glu	234	134	17	83	51
Gly	422	91	62	269	194
His	129	49	22	58	36
Ile	233	95	73	65	32
Leu	358	164	91	103	62
Lys	347	153	50	144	103
Met	73	40	15	18	13
Phe	170	73	46	51	30
Pro	176	38	19	119	79
Ser	367	107	54	206	155
Thr	278	87	65	126	79
Trp	78	32	21	25	22
Tyr	184	48	53	83	62
Val	357	144	119	94	53
总数	4741	1798	930	2013	1400
频率 c		$f_\alpha, 0.38$	$f_\beta, 0.20$	$f_c, 0.42$	$f_t, 0.30$

- a. α, β, C, T , 分别为螺旋、 β -结构、无规卷曲和 β -回折。
 b. β 回折的残基数(n_i)是单独算的,重迭的 β -回折残基只计算一次。
 c. $f_\alpha + f_\beta + f_c = 1$

$f_c = 1$, 而 β 回折是单独算出的。处于 α 或 β -结构末端的残基,也可处于回折的第一或第四位上。由图 1A 可知,由七、八个残基组成的螺旋最多;至少要四个残基才能形成一圈 α_1 螺旋。由图 1B 可知,由五、六个残基组成的 β 结构最多,有三个蛋白质的 β 结构仅由二个残基组成。图 1A 和 1B 比较而言,螺旋的残基数可多达二十几个,而 β 结构的残基数大于十二个的就很少见。可见总的看来螺旋的长度大于 β -结构的长度。其实,这也毫不足怪,螺旋是靠链内氢键维系的, β -结构则靠链间氢键维系,而链内氢键显然比链间氢键容易形成。

表 2 为由 29 个蛋白所得各氨基酸残基形成螺旋、 β 结构的构象参数。由表 1 我们可知各残基分处 α, β, C, T 中的频率,如对 Ala,

表 2 由 29 个蛋白所得形成 α -螺旋和 β -结构的
残基构象参数 P_α, P_β

α -残基	P_α	螺旋判别	β -残基	P_β	β -判别
Glu ⁻	1.51	H _{α}	Val	1.70	H _{β}
Met	1.45	H _{α}	Ile	1.60	H _{β}
Ala	1.42	H _{α}	Tyr	1.47	H _{β}
Leu	1.21	H _{α}	Phe	1.38	h _{β}
Lys ⁺	1.16	h _{α}	Trp	1.37	h _{β}
Phe	1.13	h _{α}	Leu	1.30	h _{β}
Gln	1.11	h _{α}	Cys	1.19	h _{β}
Trp	1.08	h _{α}	Thr	1.19	h _{β}
Ile	1.08	h _{α}	Gln	1.10	h _{β}
Val	1.06	h _{α}	Met	1.05	h _{β}
Asp ⁻	1.01	l _{α}	Arg ⁺	0.93	i _{β}
His ⁺	1.00	l _{α}	Asn	0.89	i _{β}
Arg ⁺	0.98	i _{α}	His ⁺	0.87	i _{β}
Thr	0.83	i _{α}	Ala	0.83	i _{β}
Ser	0.77	i _{α}	Ser	0.75	b _{β}
Cys	0.70	i _{α}	Gly	0.75	b _{β}
Tyr	0.69	b _{α}	Lys ⁺	0.74	b _{β}
Asn	0.67	b _{α}	Pro	0.55	B _{β}
Pro	0.57	B _{α}	Asp ⁻	0.54	B _{β}
Gly	0.57	B _{α}	Glu ⁻	0.37	B _{β}

H _{α} , H _{β} 为最易形成 α 或 β 结构的残基; h _{α} , h _{β} 为易形成 α 或 β 结构的残基; l _{α} , l _{β} 为形成 α 或 β 结构趋势较弱的残基; i _{α} , i _{β} 为形成 α 或 β 结构无倾向性的残基; b _{α} , b _{β} 为破坏 α, β -结构形成的残基; $\beta_\alpha, \beta_\beta$ 是最易破坏 α, β 结构形成的残基。

$f_\alpha = 234/434 = 0.539$, $f_\beta = 71/434 = 0.164$;
 $f_c = 85/434 = 0.196$ 。而表 2 中各残基形成 α, β, C, T 的构象参数分别定义为 $P_\alpha = f_\alpha / \langle f_\alpha \rangle$, $P_\beta = f_\beta / \langle f_\beta \rangle$, $P_c = f_c / \langle f_c \rangle$, $P_t = f_t / \langle f_t \rangle$ 。其中 $\langle f_\alpha \rangle \langle f_\beta \rangle \langle f_c \rangle \langle f_t \rangle$ 分别为所有残基在 α, β, C, T 中出现的平均频率,则对 Ala, $P_\alpha = 0.539/0.378 = 1.42$, $P_\beta = 0.164/0.197 = 0.82$; $P_c = 0.196/0.295 = 0.66$ 。由表 2 可知,最易形成 α 螺旋的残基是 Glu⁻, Met, Ala, Leu; 最易形成 β 结构的残基是 Val, Ile, Tyr; 易使 α 螺旋中断的残基是 Gly, Pro, Asn, Tyr; 最易破坏 β 结构的是 Glu⁻, 所有五个荷电残基 P_β 均小于 1.0, 表明它们不利于形成 β 结构。

表 2 所列构象参数虽指明了各残基在蛋白质中形成 α, β 等二级结构的趋势,但没有指明它们在各个二级结构区中的位置选择。表 3 表 4 列出的 29 个蛋白中,处于 α 或 β 结构边缘的

表 3 由 29 个蛋白所得螺旋边缘残基的构象参数

$P_{\alpha N}$		$P_{\alpha C}$		$P_{\beta \alpha N}$		$P_{\beta \alpha C}$	
Glu(-)	2.44	Lys(+)	1.83	Ser	1.55	His(+)	1.86
Asp(-)	2.02	His(+)	1.77	Asn	1.42	Asn	1.64
Pro	2.01	Met	1.57	Gly	1.41	Gly	1.64
Trp	1.47	Val	1.25	His(+)	1.22	Pro	1.58
Ala	1.29	Arg(+)	1.25	Pro	1.10	Lys(+)	1.49
Gln	1.22	Glu(-)	1.24	Thr	1.09	Arg(+)	1.24
Thr	1.08	Gln	1.22	Glu(-)	1.04	Asp(-)	1.06
Asn	0.81	Ala	1.20	Lys(+)	1.01	Phe	1.04
Gly	0.76	Leu	1.13	Tyr	0.99	Tyr	0.96
Ser	0.74	Cys	1.11	Asp(-)	0.98	Cys	0.94
His(+)	0.73	Phe	1.10	Phe	0.93	Ser	0.93
Met	0.71	Ile	0.98	Leu	0.85	Ile	0.87
Tyr	0.68	Ser	0.96	Met	0.83	Thr	0.86
Ile	0.67	Thr	0.75	Ile	0.78	Leu	0.84
Cys	0.66	Tyr	0.73	Gln	0.75	Gln	0.70
Lys(+)	0.66	Asp(-)	0.61	Val	0.75	Glu(-)	0.59
Phe	0.61	Asn	0.59	Ala	0.70	Ala	0.52
Val	0.61	Gly	0.42	Cys	0.65	Met	0.52
Leu	0.58	Trp	0.40	Trp	0.62	Val	0.32
Arg(+)	0.44	Pro	0.00	Arg(+)	0.34	Trp	0.16

$P_{\alpha N}$ 、 $P_{\alpha C}$ 分别为在 α 螺旋的 N 端或 C 端三个位置上出现的残基归一化频率。即

$$P_{\alpha N} = f_{\alpha N} / \langle f_{\alpha N} \rangle; P_{\alpha C} = f_{\alpha C} / \langle f_{\alpha C} \rangle。$$

$P_{\beta \alpha N}$ 、 $P_{\beta \alpha C}$ 分别为在 α 螺旋 N 端或 C 端外侧三个残基位置上出现的归一化频率。

残基构象参数，由 29 个蛋白的 152 段螺旋和 137 段 β 区，可分析出各残基处于 α 或 β 结构 N 端三个残基位置及 C 端三个残基位置的频率；再以所有 20 种残基的相应平均频率除之，则为归一化频率。由表 3 可知，负电残基 Glu^- 、 Asp^- ，倾向于在 α 螺旋 N 端三个残基位置上出现，而正电残基 Lys^+ 、 His^+ 、 Arg^+ 则经常出现于 α 螺旋 C 端的三个残基位置上。Pro 绝不出现于螺旋 C 端；它经常出现于 N 端，其频率为所有残基的平均决频率的两倍 ($P_{\alpha N} = 2$)，两个极性残基 Ser、Thr 经常出现于螺旋 N 端的外侧。由表 4 可知荷电残基不易在 β 结构两端出现。常出现于 β 结构 N 端的是 Ile、Val、Glu；而常出现于 C 端的是 Tyr、Val、Phe。Pro 和 Gly 则经常处于 β 结构的两侧。综合表 2 和表 4，可知疏水残基包括芳香残基 (Tyr、Phe、Trp) 和有较大烷基侧链的残基 (Ile、Val、Leu 以至 Cys) 都易形成 β 结构。由表 2—4，可综合分析各肽段形成 α 或 β 结构的趋势。

β -turn 或称 β 回折、 U 型回折、发夹结构等，是以 $4 \rightarrow 1$ 氢键稳定的四肽结构。它广泛存在于球蛋白中，按 29 个蛋白的 408 个 β -回折统计，处于回折中的残基高达 30%。由于 β -回折可使肽链产生 180° 的转折，它对蛋白质复杂而多样的三维折迭可能具有很关键的作用。仍以 Ala 为例，在回折的第一至第四位上出现的次数分别为 26, 33, 15 和 25，以表 1 中 Ala 的总数 434 除之，得 4 个位置上的出现频率分别为 $f_i = 0.06$, $f_{i+1} = 0.076$, $f_{i+2} = 0.035$, $f_{i+3} = 0.058$ 。而 Ala 在回折中的总数 $n_i = 85$ (见表 1) 小于 $\sum_{i=1}^4 n_i = 99$ ，因为重复的同一残基不计算两次，若 Ala_{12} 在 9—12 及 11—14 两个 β 回折中出现，则 n_i 中只计一次。同样，在第二、三位上只有 45 个 Ala，这样，Ala 在整个回折中和在其二、三位上出现的频率分别为 $f_i = 85/434 = 0.196$ 和 $f_{i_2} = 45/434 = 0.104$ ，而所有 20 种残基的平均频率分别为 $\langle f_i \rangle = 1400/$

表 4 由 29 个蛋白所得 β -结构边缘残基的构象参数

$P_{\beta N}$		$P_{\beta C}$		$P_{\# \beta N}$		$P_{\# \beta C}$	
Ile	1.94	Tyr	1.96	Asn	1.86	Pro	1.69
Val	1.69	Val	1.79	Pro	1.58	Gly	1.68
Gln	1.65	Phe	1.50	Gly	1.46	Trp	1.59
Phe	1.50	Ile	1.35	Ser	1.41	Ser	1.49
Trp	1.49	Leu	1.27	Asp(-)	1.39	Asp(-)	1.32
Met	1.43	Asn	1.21	Cys	1.34	Thr	1.16
Leu	1.30	Trp	1.19	Tyr	1.23	Asn	1.13
Thr	1.17	Cys	1.11	Lys(+)	1.09	Arg(+)	1.05
Tyr	1.07	Met	0.95	Gln	1.09	Tyr	1.01
Lys(+)	1.00	His(+)	0.90	Thr	1.09	His(+)	0.96
Arg(+)	0.90	Arg(+)	0.90	Glu(-)	0.92	Met	0.85
Cys	0.87	Asp(-)	0.85	Arg(+)	0.89	Glu(-)	0.85
Ala	0.86	Ser	0.79	His(+)	0.78	Lys(+)	0.82
Pro	0.66	Thr	0.75	Ala	0.67	Gln	0.77
Asn	0.66	Ala	0.75	Ile	0.59	Ala	0.74
Gly	0.63	Gly	0.74	Met	0.52	Val	0.59
Ser	0.63	Lys(+)	0.74	Trp	0.48	Leu	0.59
His(+)	0.54	Gln	0.65	Leu	0.46	Ile	0.53
Asp(-)	0.38	Glu(-)	0.55	Val	0.42	Cys	0.53
Glu(-)	0.35	Pro	0.40	Phe	0.30	Phe	0.44

$P_{\beta N}$ 、 $P_{\beta C}$ 分别为在 β 结构的 N 端或 C 端三个位置上出现的残基归一化频率。即

$$P_{\beta N} = f_{\beta N} / \langle f_{\beta N} \rangle; P_{\beta C} = f_{\beta C} / \langle f_{\beta C} \rangle$$

$P_{\# \beta N}$ 、 $P_{\# \beta C}$ 分别为在 β 结构 N 端或 C 端外侧三个残基位置上出现的归一化频率。

4741 = 0.295 和 $\langle f_{i_2} \rangle = 764/4741 = 0.161$, 则 Ala 在整个 β 回折中和在回折第二、三位上的归一化构象参数分别为 $P_i = f_i / \langle f_i \rangle = 0.66$ 和 $P_{i_2} = f_{i_2} / \langle f_{i_2} \rangle = 0.64$ 。由表 5 可知, Asn、Gly、Pro、Asp 和 Ser 在 β -回折的四个位置上出现频率最高, 而疏水残基频率最低。某些残基对 β -回折的四个位置上有所选择。如 Pro 常在第二位(30%), 而不常在第三位(3%); Trp 常在第四位(17%)而不常在第二位等。这些位置的选择, 无疑是与这些残基的立体化学制约有关。此外, Chou 和 Fasman 还对 β -回折相邻区域的残基作出统计^[4], 得到 β 回折前四个残基位置上出现的频率最大的残基分别是: (t-4)位, Trp、Val、Cys, (t-3)位, Cys、Trp、Ser; (t-2)位, His、Gly、Tyr、和(t-1)位, Arg、Trp、Tyr。而 β 回折后面四个位置上出现频率最大的残基分别为(t+1)位, Trp、Pro、Ile; (t+2)位, Pro、Cys、Asp; (t+3)位, Cys、Asp、Val; (t+4)位, Trp、Met、Glu、Ile。可见, 在 β 回折

前后有大量疏水残基出现, 而这些残基恰是易形成 α 或 β 的残基, 事实上, 在许多超二级结构, 如 α - α 、 β - β 、 β - α - β 、或 α - β 、 β - α 之间确有 β 回折存在; 也正是这些回折, 使肽链走向成 180° 逆转, 才得以形成这些超二级结构。反言之, 反平行 β -Sheet、 α - α 、 α - β 和 β - α 二级结构之间的相互作用, 亦可帮助稳定 β 回折。

至此, 我们已可对蛋白质中各肽段形成 α 、 β 、 T 等二级结构的趋势进行分析, 计算出某些肽段的平均构象参数 $\langle P_\alpha \rangle$ 、 $\langle P_\beta \rangle$ 、 $\langle P_i \rangle$ 或 $\langle P_{i_2} \rangle$ 等。然后, 根据 Chou 和 Fasman 总结的形成二级结构的判定规则, 作出二级结构预测。

规则一: 在天然蛋白中, 任何六个或更多的残基片段, 有 $\langle P_\alpha \rangle \geq 1.03$ 和 $\langle P_\alpha \rangle > \langle P_\beta \rangle$, 满足条件 1、2, 则可预测为形成 α 螺旋。条件 1, 在肽链序列上任何六个残基中有四个螺旋残基 (H_α 或 h_α), 则可形成螺旋的核心。而形成 α 螺旋的趋势较弱的残基 I_α 可计为 $1/2h_\alpha$, 即六个残基中若有三个 h_α 和二一个 I_α 即可形成

表5 由29个蛋白所得 β -回折的残基频率 (f) 及构象参数 (P)

i_1	i_{i+1}	i_{i+2}	i_{i+3}	P_i	P_{i_2}
Asn 0.161	Pro 0.301	Asn 0.191	Trp 0.167	Asn 1.56	Pro 2.04
Cys 0.149	Ser 0.139	Gly 0.190	Gly 0.152	Gly 1.56	Gly 1.63
Asp 0.147	Lys 0.115	Asp 0.179	Cys 0.128	Pro 1.52	Asp 1.61
His 0.140	Asp 0.110	Ser 0.125	Tyr 0.125	Asp 1.46	Asn 1.56
Ser 0.120	Thr 0.108	Cys 0.117	Ser 0.106	Ser 1.43	Ser 1.52
Pro 0.102	Arg 0.106	Tyr 0.114	Gln 0.098	Cys 1.19	Lys 1.13
Gly 0.102	Gln 0.098	Arg 0.099	Lys 0.095	Tyr 1.14	Tyr 1.08
Thr 0.086	Gly 0.085	His 0.093	Asn 0.091	Lys 1.01	Arg 1.05
Tyr 0.082	Asn 0.083	Glu 0.077	Arg 0.085	Gln 0.98	Thr 0.98
Trp 0.077	Met 0.082	Lys 0.072	Asp 0.081	Thr 0.96	Cys 0.92
Gln 0.074	Ala 0.076	Thr 0.065	Thr 0.079	Trp 0.96	Gln 0.84
Arg 0.070	Tyr 0.065	Phe 0.065	Leu 0.070	Arg 0.95	Glu 0.80
Met 0.068	Glu 0.060	Trp 0.064	Pro 0.068	His 0.95	His 0.77
Val 0.062	Cys 0.053	Gln 0.037	Phe 0.065	Glu 0.74	Ala 0.64
Leu 0.061	Val 0.048	Leu 0.036	Glu 0.064	Ala 0.66	Phe 0.62
Ala 0.060	His 0.047	Ala 0.035	Ala 0.058	Met 0.60	Met 0.51
Phe 0.059	Phe 0.041	Pro 0.034	Ile 0.056	Phe 0.60	Trp 0.48
Glu 0.056	Ile 0.034	Val 0.028	Met 0.055	Leu 0.59	Val 0.43
Lys 0.055	Leu 0.025	Met 0.014	His 0.054	Val 0.50	Leu 0.36
Ile 0.043	Trp 0.013	Ile 0.013	Val 0.053	Ile 0.47	Ile 0.29

$i_1 \dots i_{i+3}$ 分别为 β -回折的第一至第四个位置上出现的残基频率。 P_i 为在 β -回折的所有四个位置上出现的构象参数。 P_{i_2} 为在 β -回折的第二、三位上出现的构象参数。

螺旋的核心。条件2, 由于 Pro、Asp⁻、Glu⁻常出现于螺旋N端, His⁻、Lys⁺、Arg⁺常出现于螺旋C端, 若为满足条件1, 可将靠近螺旋N端的 Pro、Asp、当作 I_α , 靠近螺旋C端的 Arg 也可作 I_α , 而在螺旋N端的 Glu 还可当作 H_α , C端的 His、Lys 则当作 h_α 和 I_α 。若螺旋遇到 $\langle P_\alpha \rangle < 1$ 的由破坏螺旋形成的残基所组成的四肽段, 即遇到 $b_4, b_3i, b_2i_2, b_2ih, bi_3, bi_2h, bih_2$ 和 i_4 的肽段, 螺旋不再延伸。某些四肽, 如 hi_3 和 h_2i_2 , 其 $\langle P_\alpha \rangle < 1.0$, 但未列为破坏螺旋肽段, 因它们可允许螺旋继续延伸。(如 $bbhi$ 的 hi 可作为螺旋的N端, 但 bb 残基不参与形成螺旋的两端)。在四肽中 i, b, h 符号也包括了 I, B, H 。 $\langle P_\beta \rangle > \langle P_\alpha \rangle$ 的相邻 β 区域也可中止螺旋的延伸。

规则二, 在天然蛋白中三个或更多残基的片段, 有 $\langle P_\beta \rangle \geq 1.05$ 和 $\langle P_\beta \rangle > \langle P_\alpha \rangle$, 满足条件3、4, 则可预测为形成 β 结构。条件3, 在肽链序列上有连续三个 H_β 或 h_β , 或者在四、五个残基中有三个 H_β 或 h_β , 则可形成 β 结

构的核心。条件4, β 结构遇到 $\langle P_\beta \rangle < 1$ 的由破坏 β 结构的残基组成的四肽即中断。 $\langle P_\alpha \rangle > \langle P_\beta \rangle$ 的相邻 α 区, 也可中止 β 结构延伸, 荷电残基和 Pro 不利于形成 β 结构, 除非它们出现于四肽中, 且 $\langle P_\alpha \rangle < \langle P_\beta \rangle > 1$ 。

规则三, 同时可形成 α 和 β 结构的肽段, 若 $\langle P_\alpha \rangle > \langle P_\beta \rangle$ 则形成 α 螺旋; 若 $\langle P_\beta \rangle > \langle P_\alpha \rangle$ 则形成 β 结构。此外, 还有几点补充判别规则: 一段较长的 α 螺旋中, 它可含一段较短的 β 结构, 只要整体而言 $\langle P_\alpha \rangle > \langle P_\beta \rangle$, 且中间无包含 b_α, B_α 的四肽, 整个肽段都呈 α 螺旋。若一肽段既可形成 α 螺旋, 又可形成反平行 β 折迭, 则形成反平行 β 折迭。有时要利用形成 α 或 β 的边缘残基频率来综合判断, 详见文献[4]。与 β 回折相邻的片段, 若既可形成 α 螺旋, 亦可成 β 结构, 则只要两侧有三个 h_β 或 H_β , 即形成只有三个残基的 β 结构。

规则四, β -回折的判定方法有二个。既可简单地由 P_i 值算出四个残基的肽段平均 $\langle P_i \rangle > 1.0$ 及 $\langle P_\alpha \rangle < \langle P_i \rangle > \langle P_\beta \rangle$ 来判定; 亦可由

四肽的 $P_i = f_i \times f_{i+1} \times f_{i+2} \times f_{i+3} > 0.75 \times 10^{-4}$ 来判定形成 β 回折。

Chou 和 Fasman 对于预测方法的精度,对于形成 α 螺旋, β 结构的过程还作了许多分析,并设计了计算程序¹⁴。但仅根据表 2—5 及上述判定规则,即可对蛋白质的各肽段作出二级结构预测。将腺苷酸激酶的晶体衍射结果与十种二级结构预测结果进行比较,就预测 α 、 β 、 T 的总体结果而言,Chou 和 Fasman 的方法是令人满意的^{13,14}。

我们在研究胰岛素 A、B 链相互作用时,曾

以紫外差光谱、核磁共振氢谱¹⁰等方法,观察到 S-硫甲基型 A、B 链相混后,因次级链相互作用,使肽链构象有所调整,Tyr 残基由相对亲水的环境转移到相对疏水的环境。随后,以圆二色性方法观察了 A、B 链相互作用过程中的二级结构变化,测定了 S-硫甲基型及还原型 A、B 链的二级结构⁹。结果表明单独的 A、B 链具有一定二级结构。表 6 为我们根据 Chou 和 Fasman 的构象参数及判定规则对胰岛素 A、B 链二级结构的分析和预测。可作为一实例,看 Chou-Fasman 法的应用。

表 6 猪胰岛素 A、B 链的二级结构序列预测

肽 段	P_α	P_β	$P_{\alpha N}$	$P_{\beta N}$	$P_{\alpha C}$	$P_{\beta C}$	预 测
A ₂ -A ₇	1.03	1.19	0.67		0.75		成 α , 因 A ₄ Glu 为 B β , 不能成 β
A ₁₂ -A ₁₇	1.08	1.05	0.74		1.24		成 α , 因 A ₁₇ Glu/A ₁₄ Asp 为 B β
A ₁₃ -A ₁₆	1.05	1.29	0.58	1.30	1.13	1.27	成 β 趋势大,也可成 α
A ₅ -A ₁₀	0.84	1.17		1.65		1.35	有可能成 β
A ₅ -A ₁₁	0.87	1.17		1.65		1.11	有可能成 β
A ₅ -A ₈	0.84	1.17		1.65		0.75	有可能成 β
B ₂ -B ₆	1.01	1.17		1.25		1.27	成 β ,
B ₁ -B ₆	1.03	1.21	0.61	1.50	1.13	1.27	既可成 β , 也可能成 α
B ₁ -B ₇	0.98	1.20		1.50		1.11	成 β ,
B ₉ -B ₁₉	1.08	1.16	0.74		1.11		成 α , 因 B ₁₃ Glu 为 B β , 不能成 β
B ₁₃ -B ₁₉	0.97	1.39		1.30		1.11	成 β
B ₂₄ -B ₂₉	0.92	1.12		1.50		0.74	成 β

各符号含义均见正文。 $P_{\alpha N}P_{\alpha C}P_{\beta N}P_{\beta C}$ 均只列出最末端的残基构象参数

由表 6 可知, A₂-A₇ 及 A₁₂-A₁₇ 两肽段的序列预测均为 α 螺旋, 此结果与胰岛素的晶体结构分析结果一致^{10,11}。这两个肽段形成 β 结构的构象参数 P_β 均较高,但由于 A₄、A₁₇Glu 和 A₁₄Asp 均为 B β , 最易破坏 β 结构,故不可能形成 β 结构。此外,序列预测表明 A 链还有四个肽段有形成 β 结构的可能,即 A₁₃-A₁₆、A₅-A₈、A₅-A₁₀、A₅-A₁₁。其中 A₁₃-A₁₆ 形成 β 结构的可能最大,因其 P_β 值很高,为 1.29。而 A₁₃Leu 的 $P_{\beta N}$ 以及 A₁₁Leu 的 $P_{\beta C}$ 均远高于一般残基,故易形成 β 结构。不过,也不能排除其形成 α 的可能。因其 $P_\alpha > 1.03$,而 A₅-A₈、A₅-A₁₀、A₅-A₁₁ 三段的 P_β 均为 1.17,而 P_α 均远小于 1.03,故亦可能成 β 结构。但在胰岛素分子中, A₆-A₁₁ 为二硫键,形成链内小环。因此很难想

像,在此二十元环中又包括一段比较伸展的 β 结构。只有在全部还原的 A 链中,这三段或许有形成 β 结构的可能。我们测得 S-硫甲基型 A 链的 α 螺旋, β 结构含量为 15% 及 27%; B 链为 22% 及 23%; 两链相混两小时后 α 、 β 含量为 24% 及 26%。实际测得 A 链不仅含有部分 α 螺旋,还含有较多 β 结构成份。而胰岛素晶体分析表明, A 链只含 α 螺旋,不含 β 结构^{10,11}。这里可能既涉及晶体结构与溶液构象的差别,也涉及胰岛素分子中 A 链的肽链走向与单独的 A 链的溶液构象的区别,以至 S-硫甲基的影响。Chou 和 Fasman 的预测方法使我们可以推测,或许单独的 S-硫甲基型 A 链的 A₁₃-A₁₆ 可形成 β 结构。此外, A₁₇-A₁₉ 在晶体中为 β 回折,按序列预测计算, A₁₇-A₂₀ 确有

可能形成 β 回折,其 $\langle P_i \rangle = 1.13, P_i = 0.90 \times 10^{-4}$ 。但由于单独的A链,既无 $A_{20}-B_{19}$ 链间二硫键,也没有 $A_{21}-B_{22}$ 盐键,因此其C端难以稳定,可能实际上不能形成 β -回折。

由表5序列预测还表明 $B_2-B_6, B_{24}-B_{29}$ 可形成 β 结构, B_9-B_{19} 可形成 α 螺旋,这与2-Z_n胰岛素晶体结构报道是一致的^[10]。但序列预测法计算 B_1-B_6 肽段,表明它既可成 α 螺旋,亦可成 β 结构。由于 P_β 为1.21,大于 P_α 1.03,故一般表现为 β 结构,如2-Z_n胰岛素晶体中的情况^[10]。但按 α, β 的边缘残基分析,则 B_4, B_5, B_6 的 $P_{\alpha C}$ 分别为1.22,1.77,1.13。而 $P_{\beta C}$ 则为0.65,0.90,1.27。因此 B_1-B_6 肽段的C端残基形成 α 的可能又大于 β 。加之, P_α 1.03亦可为 α 螺旋。实际上,在6% NaCl中结晶出的4-Z_n胰岛素B链N端确实为 α 螺旋^[11]。Chou和Fasman对此也有所讨论^[3,4]。这充分表明,序列预测法确实可提出各肽段形成二级结构的各种可能性,而实际上究竟形成什么结构,还视晶体或溶液的具体状况,包括溶剂状况、二硫键断裂与否、荷电残基的状况而定。此外,B链的 P_i 分析指出,晶体结构中形成回折的 $B_{20}-B_{22}$ 和 B_7-B_9 附近, P_i 确实大于1。不过在单独的B链中,无 $A_7-B_7, A_{19}-B_{20}$ 两对二硫键,也没有 $B_{22}-A_{21}$ 盐键及 $A_{19}-B_{25}$ 氢键维系,故此两段 β -回折,特别是 $B_{20}-B_{22}$ 难以稳定。很可能在单独的B链中,这部份肽段呈无规卷曲状态(因有 B_8, B_{20} 两个Gly)。

Chou-Fasman的预测方法是由统计学得出

的经验推断方法。因此排除例外,具体事物还要具体分析;但既作为统计学的结果,则一定还包含某种客观规律,这就是各氨基酸残基的 ϕ, ψ, ω 等二面角对形成 α, β, T 等二级结构的立体化学制约性。蛋白质的氨基酸序列之所以包含着形成二级结构以至全部高级结构的信息,也正在于此。随着蛋白质晶体结构数据库的不断增大,由残基序列预测蛋白质二级结构的精度将大大提高,其应用前景将更加广阔。

参 考 文 献

- [1] Chou, P. Y., and Fasman, G. D., *Biochemistry*, **13**, 211, 1974.
- [2] Chou, P. Y., and Fasman, G. D., *ibid* **13**, 222, 1974.
- [3] Chou, P. Y., and Fasman, G. D., *Ann. Rev. Biochem.*, **47**, 251, 1978.
- [4] Chou, P. Y., and Fasman, G. D., *Adv. Enzymology*, **47**, 45, 1978.
- [5] Cantor, C. R., and Schimmel, P. R. *Biophysical Chemistry*, Part I San Francisco, W. H. Freeman and Co
- [6] Argos, P., and Palau, J. *Int. J. Peptides Protein Res.*, **19**, 380, 1982.
- [7] 华庆新、王家槐《生物化学与生物物理进展》3, 2, 1984.
- [8] Hua Qing-Xin, Qian Yan-Qiu, Tsou Chen-Lu *Biochim, Biophys. Acta*, **789**(2), 234, 1984.
- [9] 华庆新、钱雁秋、邹承鲁《中国科学》B辑,5,404,1985.
- [10] Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. C., Mercola, D. A., and Vijayan, M. *Nature*, **231**, 506, (1971).
- [11] Bentley, G., Dodson, E., Dodson, G., Hodgkin, D., Mercola, D. *Nature*, **261**, 166, 1976.

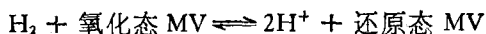
[本文于1984年7月10日收到]

细菌氢酶的结构和催化功能

朱长喜 宋鸿遇

(中国科学院 上海植物生理研究所)

早在本世纪初就发现多种生物体,主要是细菌和藻类具有释放和利用分子氢的能力,但是参与分子氢反应的酶的本质,直到30年代才得到证实。通常用下面的反应式表示:



1931年Stevenson和Stickland命名具有此催化活性的酶为氢酶。

自1974年Chen和Mortenson在巴氏梭