

文章编号 :1009-038X(2002)05-0464-04

蛋白质二硫键的分布特征

宋江宁, 李炜疆

(江南大学 工业生物技术教育部重点实验室 江苏 无锡 214036)

摘 要 :以 SLCTBASE 数据库中包含二硫键的 215 条多肽链作为研究对象,统计分析二硫键的分布特征。结果表明,二硫键的分布很不均匀,主要在序列距离较近的 2 个半胱氨酸之间形成,大部分二硫键是在序列距离小于 70 个氨基酸的地方形成,并且对某些序列距离有着强烈的偏好。这一结果对蛋白质工程中引入人工二硫键、进行分子生物学实验设计具有一定参考意义。

关键词 :半胱氨酸 ;二硫键 ;蛋白质折叠 ;统计分析

中图分类号 :Q 61

文献标识码 :A

Analysis of Sequence Features of Disulfide-Bonds in Proteins

SONG Jiang-ning, LI Wei-jiang

(The Key Laboratory of Industrial Biotechnology, Ministry of Education, Southern Yangtze University, Wuxi 214036, China)

Abstract :Based on the sampling database SLCTBASE with 215 protein polypeptide chains, we statistically analyzed the distribution features of sequence distance and reduced distance of the disulfide bonds. The results suggest that disulfide bonds are mainly formed between the two cysteines with close sequence distances and show preference for some certain sequence distance. The results will be helpful for introducing S-S bonds in protein engineering and molecular biology experiments.

Key words :cysteine ;disulfide bond ;protein folding ;statistical analysis

二硫键是一类在蛋白质多肽链的两个半胱氨酸(Cys)之间形成的共价键,这两个半胱氨酸在氨基酸序列上的分布可以相距较近,也可以较远。二硫键可以形成在蛋白质的同一个亚基内,也可以在不同亚基之间产生。通常将形成二硫键的两个 Cys 的状态称为氧化态,没有形成二硫键的 Cys 的状态则称为还原态,或者硫醇态。

二硫键的形成对于稳定蛋白质的空间结构和保持其活性功能具有极其重要的影响,它的错误配对是影响蛋白质多肽链正确折叠的重要原因。二硫键的形成是蛋白质折叠过程中的重要步骤,其形成

动力学影响着蛋白质折叠的速率和途径^[1]。二硫键除了对于单分子折叠反应具有动力学和热力学上的影响外,还可以调节细胞内环境分子间的接触,从而影响蛋白质的分泌情况。研究形成二硫键的蛋白质序列特征,找出与二硫键形成有关联的某些结构信息,对于蛋白质工程和人工药物分子设计有着积极而重要的意义。

氨基酸序列在蛋白质多肽链折叠后,序列距离远的两个氨基酸在空间距离上不一定也相距较远,例如形成二硫键的两个 Cys,它们在空间距离上离得很近,但这是否与序列距离有关,形成二硫键的

收稿日期 2002-04-24; 修订日期 2002-06-22.

基金项目 :工业生物技术教育部重点实验室访问学者基金资助.

作者简介 :宋江宁(1978-),男,山东莱阳人,发酵工程硕士研究生.

Cys 在氨基酸序列上是否是随机分布的呢？

目前主要有 3 种预测 Cys 的氧化态和硫醇态的方法 :第一种是运用神经网络识别 Cys 的状态^[2] ;第二种是对于 Cys 周围序列中氨基酸出现的频率进行统计分析^[3] ;第三种是利用多序列联配(multiple sequence alignment)预测 Cys 残基的状态(氧化态或者硫醇态)^[4] . Muskal 等人运用神经网络方法对一个较小的独立数据集进行预测 ,结果达到 81 % 的准确率^[2] . Fiser 等人用统计方法对一个比前者大 4 倍的数据集进行测试 ,预测准确率也可以达到 71 %^[3] . 2000 年 Fiser 和 Simon 利用多序列联配预测氧化态和硫醇态 Cys ,准确率达到 82 % 以上^[4] .

作者在更大的高精度蛋白质空间结构数据库 SLCTBASE^[5] 基础上 ,对于库中含有的蛋白质序列进行了统计检验 ,对 Cys 附近序列(包括含有氧化态和硫醇态的 Cys 氨基酸序列)进行统计分析 ,以研究 Cys 邻近序列是否影响二硫键的形成、二硫键在蛋白质多肽链上所处的位置以及形成二硫键的 Cys 之间的序列间隔等问题 . 与以上预测方法所不同的是 ,实验主要侧重于发掘蛋白质一维氨基酸序列中所隐藏的二硫键形成的二级结构信息 . 在蛋白质的二级结构中 ,两个氧化态 Cys 由于相互之间的空间接触而形成二硫键 ,它基本上属于一种空间结

构特征 . 预测 Cys 的存在状态也可以称之为蛋白质结构的二维预测 . 二硫键和局部序列是否有关 ? 如果通过统计分析 ,找出二硫键周围的某些序列特征 ,就可以利用这些信息预测蛋白质序列中 Cys 的状态 ,知道是哪两个 Cys 之间形成了二硫键 ,这样就可以不通过多序列联配或者神经网络的预测方法对二硫键进行预测 .

1 材料与方法

选择一个高精度非同源的蛋白质空间结构样板库 SLCTBASE^[5] (此库的蛋白质结构测定方法均为 X 射线衍射 ,测量精度均好于 2.5 Å)作为研究的样本数据库 .

选取含有肽链内二硫键的结构数据进行研究 ,计有 215 条蛋白质多肽链 ,578 条链内二硫键 ,总的氨基酸残基数为 49 612 .

1TFE 形成二硫键的两个 Cys 的氨基酸距离为 0 ,这可能是因为结构数据错误导致 ;1OBR AAAHA 的两个 Cys 距离为 1 ,这意味着紧邻的两个 Cys 形成了二硫键 ,怀疑数据可能有误 ,所以这 3 个 PDB 结构数据不使用 .

表 1 列出了 SLCTBASE 数据库中所有含链内二硫键的 PDB 文件 .

表 1 形成肽链内二硫键的 PDB 文件
Tab.1 The PDB Files of disulfide bond within the same peptide chain

PDB							
153L	1BGC	1EPTC	1HYP	1MHLC	1SVB	2AYH	2VPFA
1A6WL	1BGP	1ESC	1IAE	1MPP	1TABE	2BBKH	2WEA
1A6WH	1BHP	1EXTA	1IDK	1MUP	1TABI	2BBKL	3CD4
1AAZA	1BNDA	1EZM	1ILR1	1MZM	1TCA	2CBP	3EBX
1ABRB	1BNDB	1FLEE	1IMBA	1NEU	1TDE	2CTC	3FRUA
1AC5	1BOVA	1FLEI	1IVYA	1NNC	1TF4A	2DNJA	3FRUB
1AGQA	1BPI	1FUS	1JER	1NOYA	1TGSZ	2ENG	3GRS
1AHO	1BTL	1FVKA	1JETA	1NSCA	1TGSI	2ERL	3LADA
1AIR	1CELA	1FXD	1JFRA	1NWPA	1TGXA	2GMFA	3LZT
1AISA	1CEX	1G3P	1JMCA	1ONC	1THG	2HLCA	3PTE
1ALKA	1CFB	1GAI	1JPC	1OVA	1THV	2ILK	3SEB
1ALU	1CGT	1GAL	1KLO	1PGS	1THD	2LIV	3TGL
1AMP	1CNSA	1GEN	1KPTA	1POA	1TML	2MCM	4AAHB
1AOCA	1CNV	1GOF	1KSIA	1POC	1TN3	2MPRA	4HTCH
1AOZA	1CPO	1HCGA	1KTE	1PPN	1TVDA	2MSBA	4HTCI
1APA	1CTN	1HCGB	1KUH	1PYTB	1UKZ	2MTAH	5PTP
1APYA	1CVL	1HIAA	1KVEA	1PYTC	1UMAH	2MTAL	7RSA

续表 1

PDB							
1APYB	1DANL	1HIAB	1KVEB	1PYTD	1VCAA	2OVO	8FABA
1AQB	1DANH	1HIAI	1LBEA	1RCB	1VMOA	2PKAA	8FABB
1AQZA	1DANT	1HOE	1LBU	1RFS	1WBA	2PKAB	
1ARB	1DANU	1HPLA	1LIT	1RGEA	1WHTA	2PSPA	
1ARU	1DDT	1HSBA	1LKI	1RIE	1XJO	2RHE	
1ATLA	1DPE	1HSBB	1LPBA	1RMG	1XSOA	2SAS	
1AU1B	1EAGA	1HSSA	1LPBB	1SFP	1YAIA	2SGA	
1AUK	1ECEA	1HTRB	1LST	1SMD	1ZXQ	2SICI	
1AV4	1ECY	1HUCA	1LT5D	1SMNA	2AAA	2SIL	
1BEB A	1EDMB	1HUCB	1LTSD	1SMPI	2ACK	2TGI	
1BEO	1EPTA	1HXN	1MHLA	1SRA	2AMG	2TRXA	

2 结果与讨论

为简便起见,将形成二硫键的两个氧化态 Cys 的序列距离定义为二硫键的跨度,用它们之间相距的氨基酸个数表示.

计算出二硫键的跨度,如图 1 所示.横坐标表示二硫键的跨度,纵坐标表示相应跨度处的二硫键数目.从图上可以看出,形成二硫键的两个 Cys 的序列距离不是均匀分布的,显然地存在着某些强烈偏好.

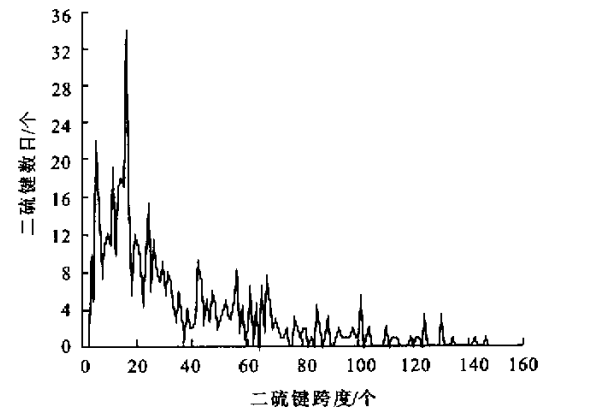


图 1 形成肽链内二硫键的 2 个氧化态 Cys 之间的序列距离

Fig. 1 Distribution of sequence distance between S-S bonded cysteines

在使用的蛋白质结构数据中,二硫键的跨度从很小到很大,但分布并不是均匀的.序列距离小于 70 的有 495 个 PDB 结构数据,这些二硫键数目总共占有二硫键数目的比例约为 86%(495/578).为了便于观察,图 1 只标出了序列距离在 150 范围以内的二硫键的形成情况.

图 1 中出现的几个峰值远高于平均数.最高的一个峰出现在序列距离为 16 处,在此位置形成二硫键的数目很多,有 36 个;但是紧挨着这个序列距离,14,15 或者 17,18 等处,二硫键的形成数目都要低很多,这已经超出了统计涨落的范围.另一个比较偏好的位置是在序列距离 5 处.图 1 的峰形都是锯齿状,而非缓慢上升或下降的形状.二硫键形成高峰紧邻着低谷,究竟是何原因?这一显著的差异究竟是由于样本选择引起的,还是由于蛋白质分子内部结构化学的原因,这有待进一步研究.

由于没有明显的几何理由,说明二硫键在形成时对某一些位置强烈偏好而对于另一些位置不偏好.这当中必然有着深层次的原因,我们或许可以从蛋白质结构特点、蛋白质化学结构以及蛋白质折叠动力学方面去找合理解释.但无论如何,这一发现对于蛋白质工程而言,有着特殊而重要的意义.如果需要引入人工二硫键,就可以优先考虑在序列距离小于 70 的范围内进行搜索,选择那些对二硫键形成有着较强偏好的位点处进行定点诱变,以避免实验的盲目性,提高实验的成功率.

由于二硫键的跨度分布不均匀,对于短的氨基酸序列,两个氧化态的 Cys 相距 50 个氨基酸,几乎是序列 N 端与 C 端形成闭合的环,而对于很长的序列,它们的距离则看作很接近.因此,引入约化距离表示两个氧化态 Cys 的相对距离,即两个氧化态 Cys 的序列距离除以它们所在的氨基酸序列的总长度.计算出来的约化距离分布见图 2.

由图 2 较易看出,形成二硫键的两个氧化态 Cys 的约化距离一般都不长,都集中在 50% 的范围内,5% 处出现一个高峰,表明二硫键比较倾向于在此约化距离处形成.

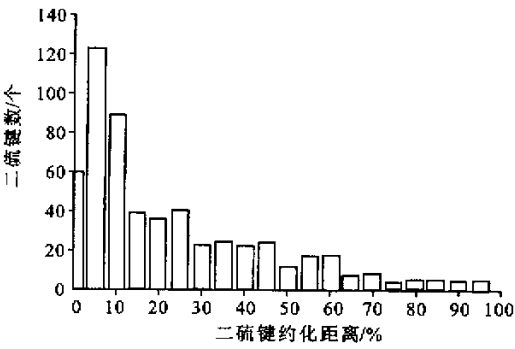


图 2 形成二硫键的两个氧化态 Cys 的约化距离
Fig.2 Distribution of reduced sequence of the oxidized Cys

如果形成二硫键的两个氧化态 Cys 之间的序列距离小于序列总长度的一半,并且这两个 Cys 都出现在序列的前面一半(或者后面一半),我们就以最靠近序列 N 端(或者 C 端)的那个氧化态 Cys 的位置作为代表此二硫键在序列中的位置,据此统计二硫键倾向于在序列前半部出现还是在后半部出现。但是如果两个氧化态 Cys 之间的序列距离大于序列总长度的一半,由于这两个氧化态 Cys 的序列距离太大,跨度太大,这样的统计结果实际上没有什么具体意义。作者排除了这种情况的二硫键。统计结果见图 3。

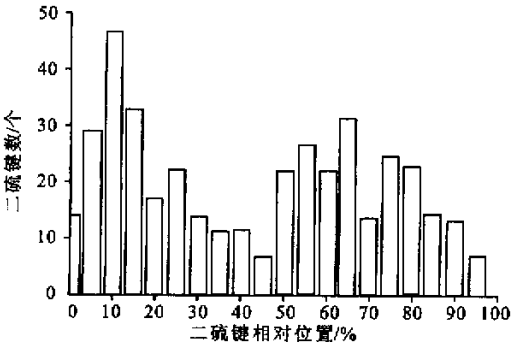


图 3 二硫键在序列中的出现位置
Fig.3 Relative position of disulfide bond in the sequence

参考文献：

[1] WITTRUP K D. Disulfide bond formation and eukaryotic secretory productivity[J]. *Current Opinion in Biotechnology* ,1995 , 6 203-208.

[2] MUSKAL S ,HOLBROOK S ,KIM S. Prediction of the disulfide-bonding state of cysteine in protein[J]. *Protein Engineering* , 1990 3 667-672.

[3] FISER A ,CSERZŐ M ,TÜDÖS E *et al.* Different sequence environments of cysteines and half cystines in proteins :Application to predict disulfide forming residues[J]. *FEBS Letters* ,1992 302 :117-120.

[4] FISER A ,SIMON I. Predicting the oxidation state of cysteines by multiple sequence alignment[J]. *Bioinformatics* ,2000 ,16 (3) 251-256.

[5] 李炜疆 ,张颖. 蛋白质结构样板库的构建及其总体特征[J]. *内蒙古大学学报* ,2000 31(5) :471-478.

[6] ZHOU H ,ZHOU Y. Folding rate prediction using total contact distance[J]. *Biophys J* 2002 82(1) :458-463.

进一步统计看出,大约有 56% 的二硫键是在靠近 N 端序列的那部分形成的,只有 44% 的二硫键是在靠近 C 端一侧形成。另外,在序列长度的前 1/4 处形成的二硫键的数目占到了二硫键总数目的 40%,这说明二硫键更倾向于在序列前半段形成。这是因为在蛋白质翻译过程中,新生肽链在延伸时如果刚开始就能够形成正确的二硫键配对,由于二硫键具有保持蛋白质构象的稳定性,那么以后生成的蛋白质多肽链将更加容易延伸和正确折叠,这样可以减少错误折叠。反之,如果大多数二硫键不是在新生肽链刚开始延伸阶段形成,而在以后的蛋白质延伸和后修饰阶段将会有很多种形成二硫键的可能性。这样对于保持蛋白质二级结构的稳定和减少错误折叠出现的频率都很不利。Zhou 等人的研究也表明,二硫键的跨度只有在较大时才会对蛋白质折叠速率产生重要影响^[6]。跨度小的二硫键对蛋白质折叠影响不大,而大跨度的二硫键既然对于蛋白质折叠速率有影响,那么对于蛋白质的折叠过程和其高级结构的形成与稳定性也会有影响。从分布上,大跨度二硫键不利于蛋白质保持其结构的稳定性,如果引入,会破坏蛋白质折叠结构。因此,在实际进行人工蛋白质分子设计时,就需要考虑引入小跨度的二硫键。

3 结 论

- 1) 二硫键的分布很不均匀,对某些序列距离有着强烈的偏好,大部分二硫键都是在序列距离小于 70 的地方形成的;
- 2) 大部分形成二硫键的两个氧化态 Cys 的约化距离都不长,多数小于 50%;
- 3) 相对而言,二硫键更倾向于在氨基酸序列的前半段出现,这在蛋白质翻译过程中对于保证蛋白质新生肽链顺利延伸合成和减少发生错误折叠是有积极意义的。