Hypothesis

Structural classification of CDR-H3 in antibodies

Hiroki Shirai, Akinori Kidera[1], Haruki Nakamura*

Department of Bioinformatics, Biomolecular Engineering Research Institute, 6-2-3 Furuedai, Suita, Osaka 565, Japan

Received 16 September 1996

Abstract Large varieties in the lengths and the amino acid sequences of the third complementarity determining region of the antibody heavy chain (CDR-H3) have made it difficult to establish a relationship between the sequences and the tertiary structures, in contrast to the other CDRs, which are classified by their canonical structures. A total of 55 CDR-H3 segments from well determined crystal structures were analyzed, and we have derived several remarkable rules, which could partly govern the CDR-H3 conformation dependence on the sequence. Since the rules are physically reasonable, they are expected to be applicable to structural modeling and design of antibodies.

Key words: Antibody; Complementarity determining region; Loop modeling; Structure classification

1. Introduction

The antibody combining site is composed of six complementarity determining regions (CDRs) [1–3]. The structural repertoires and the relationships between the amino acid sequences and the tertiary structures have been extensively studied to reveal the importance of the typical loop structures, which are canonical structures, in the three CDR segments in the light chain (L1, L2, and L3) and the first two CDR segments in the heavy chain (H1 and H2) [4–10]. Individual canonical structures are frequently observed in antibody crystal structures, and they are classified depending upon the segment lengths and the positions of specific amino acid residues in the segments.

Based upon the canonical structures, many structural models have been built in order to understand the molecular basis of antigen recognition, by various knowledge-based approaches using empirical rules [11–15], by conformation search calculations [16–18], and by combining both empirical and computational approaches [19–23].

However, the third CDR of the heavy chain (H3) has large variety in its length and amino acid sequence, and so no canonical structures have ever been established [5,9,24]. It is also recognized that the CDR-H3 has a crucial role in mediating the individual recognition of antigens [25,26], sometimes by changing its conformation upon antigen binding [27–32]. Therefore, much more attention should be paid to the structural characteristics of CDR-H3. If the CDR-H3 local structures are classified, even for only some of the segments, and the relationship between the amino acid sequences and the conformations is determined, even for specific sequences with particular features, such limited information would still be very helpful to build reliable three-dimensional (3D) models of antibodies from only the amino acid sequences.

Here, we propose a novel classification of the CDR-H3 structures, and reveal several remarkable relationships between their sequences and the loop conformations, from an analysis of the well determined crystal structures of antibodies.

2. Methods

The atomic coordinates of the antibodies were obtained from the Brookhaven Protein Data Bank (PDB; release #76) [33]. The structures of the highest resolution were first selected as representatives for the free and the complexed antibodies. Then, a total of 55 CDR-H3 segment structures with resolutions equal or better than 2.8 Å were prepared from these representatives (Table 1). In particular, the well determined segments with temperature factors less than 30 Å^2 are marked by bold letters.

The definition of the CDR-H3 segments, which are enclosed by parentheses in Table 1, follows from reference [3]. The lengths n of the CDR-H3 segments are quite variable, so the segment residues were re-numbered from 1 to n, corresponding to the conventional residue numbers from 95 to 102 of the heavy chain, as indicated in Table 1.

For a survey of the characteristics of the CDR-H3 sequences with lengths from 5 to 17, 2263 sequences of CDR-H3 were obtained from Kabat Database (Aug. 14, 1996) [3]. Here, identical CDR-H3 sequences were removed to prevent redundancy.

The structures were displayed on a computer graphics system (Indy-XZ; Silicon Graphics Inc.), and were observed with the graphics program INSIGHT-II (Molecular Simulations Inc.).

3. Kinked and extended base conformations

From a naive comparison of the amino acid sequences of the CDR-H3 segments, the first residue and the last three residues ($n-2$, $n-1$, and n) seem to form a 'common motif' that is fairly well conserved, in contrast to the intervening loop region, which exhibits great variety depending on the associated antigens [4,12]. However, when one compares the 3D structures of the CDR-H3 segments, it is obvious that these 'base' regions of the CDR-H3 have structural variety in the backbones as well, even if the segment lengths are the same and the amino acid sequences are similar. Mas et al. [21]

*Corresponding author. Fax: (81) (6) 872-8219.

[1]*Present address:* Graduate School of Science, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606-01, Japan.

Abbreviations: CDR, complementarity determining region; 3D, three-dimensional

PII S0014-5793(96)01252-5

pointed out that the backbone conformations of the bases of CDR-H3 fall into two classes: kinked and extended. In the former, kinked class, the segment is well-packed with a bulge around the $(n-1)$th residue of the CDR-H3. In the latter, extended class, the segment has no such deformation, and forms an extended β-hairpin conformation.

Therefore, we first investigated the structural features of the bases of CDR-H3. Instead of an intuitive graphics observation, the dihedral angles formed by the four Cα atoms at the $(n-2)$th, the $(n-1)$th, the nth, and the $(n+1)$th residues were calculated for the individual CDR-H3 segments, and they are plotted in Fig. 1. There are two distinct clusters, which correspond to the kinked and the extended classes (Fig. 2a,b), respectively. Thus, the current base structures are well classified into the above two classes, although many more antibody structures were investigated than those studied by Mas et al. [21]

Secondly, the relationships between the amino acid sequences and the above two structural classes were investigated. It has been pointed out that a salt bridge between

Table 1
CDR-H3 structures

n[1]	antibody	PDB code[2] free/complex		sequence[3]	resolution (Å) free/complex		antigen
				FR3 (CDR-H3) FR4			
				1 n			
5	CHA255		/1IND	**CAS**----------**(HRFVH)W**	2.2		metal chelate
5	50.1	1GGC	/1GGI	**CAQ**----------**(EGYIY)W**	2.8	/2.8	peptide
7	JEL103	1MRC	/1MRD	CAN----------(LRGYFDY)W	2.4	/2.3	poly(rI)
7	4-4-20		/1FLR	CTG----------(SYYGMDY)W		/1.85	fluorescein
7	NC6.8	1CGS	/2CGR	**CTR**----------**(GYSSMDY)W**	2.6	/2.2	NC174
7	HyHEL5		/1BQL	**CLH**----------**(GNYDFDG)W**		/2.65	lysozyme
7	TE33		/1TET	**CAR**----------**(RSWYFDV)W**		/2.3	peptide
7	D44.1	1MLB	/1MLC	**CAR**----------**(GDGNYGY)W**	2.1	/2.1	lysozyme
8	D1.3	1VFA	/1VFB	**CAR**---------**(ERDYRLDY)W**	1.8	/1.8	lysozyme
8	YST9.1	1MAM		**CTR**---------**(DPYGPAAY)W**	2.45		
8	PLG	1PLG		**CAR**---------**(GGKFAMDY)W**	2.8		
9	D11.15		/1JHL	**CAR**--------**(DDNYGAMDY)W**		/2.4	lysozyme
9	J539		/2FBJ	**CAR**--------**(LHYYGYNAY)W**		/1.95	galactan
9	SE155-4		/1MFA	**CTR**--------**(GGHGYYGDY)W**		/1.7	saccharide
9	JEL142		/1JEL	**CAR**--------**(VMGEQYFDV)W**		/2.8	HPr
9	NEW	7FAB		**CAR**--------**(NLIAGGIDV)W**	2.0		
9	8F5	1BBD		**CDG**--------**(YYSYYDMDY)W**	2.8		
10	B13I2	1IGF	/2IGF	**CTR**-------**(YSSDPFYFDY)W**	2.8	/2.8	peptide
10	17-IA	1FOR		**CAR**-------**(SGNYPYAMDY)W**	2.75		
10	R6.5	1RMF		**CAR**-------**(GGWLLLSFDY)W**	2.8		
10	DB3	1DBA	/1DBJ	**CTR**-------**(GDYVNWYFDV)W**	2.8	/2.7	steroid
10	BV04-01	1NBV	/1CBV	**CVR**-------**(DQTGTAWFAY)W**	2.0	/2.66	ssDNA $d(pT)_3$
10	17E8		/1EAP	**CKR**-------**(SYYGSSYVDY)W**		/2.5	Nle phosphonate
10	26-10	1IGI	/1IGJ	**CAG**-------**(SSGNKWAMDY)W**	2.7	/2.5	digoxin
11	4D5ver4	1FVD		**CSR**------**(WGGDGFYAMDV)W**	2.5		
11	4D5ver7	1FVE		**CSR**------**(WGGDGFYAMDY)W**	2.7		
11	4D5ver8	1FVC		**CSR**------(WGGDGFYAMDY)W	2.2		
11	17/9	1HIL	/1IFH	**CAR**------**(RERYDENGFAY)W**	2.0	/2.8	peptide
11	26/9		/1FRG	**CAR**------**(RERYDEKGFAY)W**		/2.8	peptide
11	McPC603	1MCP		**CAR**------**(NYYGSTWYFDV)W**	2.7		
11	NC41		/1NCA	**CAR**------**(GEDNFGSLSDY)W**		/2.5	neuraminidase
11	MOPC21	1IGC		CAR------(WGNYPYYAMDY)W	2.6		
12	36-71	6FAB		**CAR**-----**(SEYYGGSYKFDY)W**	1.9		
12	HIL	8FAB		**CAR**-----**(DPDILTAFSFDY)W**	1.8		
12	POT	1IGM		**CAK**-----**(HRVSYVLTGFDS)W**	2.3		
13	L5MK161	1LMK		**CAR**----**(GEDYYAYWYVLDY)W**	2.6		
13	NC10		/1NMB	**CAR**----**(SGGSYRYDGGFDY)W**		/2.5	neuraminidase
13	40-50		/1IBG	**CAR**----**(FRFASYYDYAVDY)W**		/2.7	ouabain
14	HC19	1GIG		**CAR**---**(DFYDYDVFYYAMDY)W**	2.3		
15	H52	1FGV		**CAR**--**(WRGLNYGFDVRYFDV)W**	1.9		
15	R19.9	1FAI		**CAR**--(SFYGGSDLAVYY**FDS**)W	2.7		
16	OPG2	1OPG		**CTR**-(HPFYRYDGGNYYAMDH)W	2.0		
17	KOL	2FB4		**CAR(DGGHGFCSSASCFGPDY)W**	1.9		
17	3D6	1DFB		**CVK(GRDYYDSGGYFTVAFDI)W**	2.7		
17	R45-45-11		/1IKF	**CTR(HTLYDTLYGNYPVWFAD)W**		/2.5	cyclosporin A

[1]The length of the CDR-H3 segments defined by [3].
[2]Bold letters indicate that the coordinates of all the backbone atoms in the CDR-H3 were well determined with temperature factors less than 30 Å^2.
[3]One-letter amino acid codes are used for sequences of the CDR-H3 from 1 to n surrounded by parentheses. FR3 and FR4 are the preceding and the proceeding frameworks defined by [3]. The nth residues are aligned at the same position in the CDR-H3. Bold letters indicate that temperature factors of all the backbone atoms of the residue were less than 30 Å^2. Underlined letters indicate that temperature factors of any backbone atoms of the residue were larger than 30 Å^2 in the complexed form, only when both the free and complexed forms are registered.

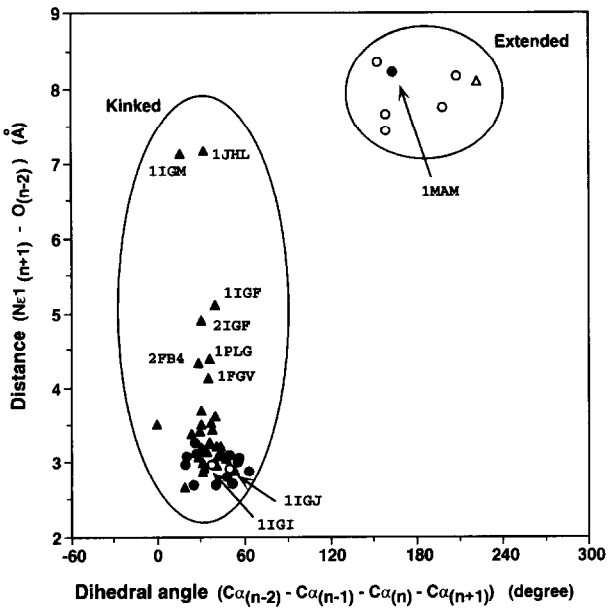

Fig. 1. Distribution of the dihedral angles formed by the four Cα atoms at the $(n-2)$th, the $(n-1)$th, the nth, and the $(n+1)$th residues, together with the distances between the backbone oxygen of the $(n-2)$th residue and the Nε1 atom of the $(n+1)$th Trp side chain. Two distinct clusters discriminated by the dihedral angles correspond to the kinked (0°–60°) and the extended ($\approx 180°$) base structures shown in Fig. 2. The symbols correspond to classification rule (i). For rule (i-a) (●), where the $(n-1)$th residue is not Asp. For rule (i-b) (○), where the $(n-1)$th residue is Asp, and the 0th residue is neither Arg nor Lys. For rule (i-c) (▲), where the $(n-1)$th residue is Asp, the 0th residue is Arg or Lys, and the (-1)th residue is neither Arg nor Lys. For rule (i-d) (△), where the $(n-1)$th residue is Asp, and both the (-1)th and the 0th residues are Arg or Lys. Arrows indicate the three exceptions to the rules. Several kinked bases without a hydrogen bond between the backbone oxygen of the $(n-2)$th residue and the $(n+1)$th Trp side chain are also indicated with the PDB codes.

the 0th Arg and the $(n-1)$th Asp is frequently observed in different antibodies, and it may contribute to stabilizing the base conformation [4,13,21,34]. This salt bridge is possible only when the backbone of the base is kinked.

However, the salt bridge itself is not an absolute condition for the formation of the kinked base, because many exceptions were observed. In fact, the CDR-H3s in CHA255, 50.1, D44.1, J539, BV04-01, 26-10, 17/9, 26/9, and R45-45-11 form the kinked base structures without the 0th basic residue or the $(n-1)$th Asp. In contrast, the base in the CDR-H3 of 17E8 forms the extended structure without a salt bridge, although the 0th and the $(n-1)$th residues are Arg and Asp, respectively.

In most kinked base structures, we have found a conserved hydrogen bond between the backbone carbonyl oxygen of the $(n-2)$th residue and the Nε1 atom of the $(n+1)$th Trp side chain (Figs. 1 and 2a). This Trp residue is very well conserved as the first residue of the proceeding framework. In contrast, most of the extended bases have alternating hydrogen bonds between the side-chain carboxyl of the $(n-1)$th Asp and the Nε1 atom of the $(n+1)$th Trp (Fig. 2b). Therefore, the side chain of the $(n-1)$th Asp is considered to have the role of switching between the kinked and the extended bases. As shown in Fig. 1, the common hydrogen bond between the backbone oxygen of the $(n-2)$th residue and the $(n+1)$th Trp side chain is sometimes lacking in several kinked bases, in all of which the salt bridges between the 0th Arg and the $(n-1)$th Asp are formed. Consequently, in the kinked bases, either the aforementioned characteristic hydrogen bond or the salt bridge is always observed.

Both the (-1)th and the 0th positions are sometimes occupied simultaneously by basic residues. In the case of 17E8, in which Lys and Arg residues occupy in the (-1)th and the 0th positions, respectively, a new salt bridge is formed between the side chain of the (-1)th Lys and the $(n-1)$th Asp, and the base structure is extended.

In conclusion, the relationship between the amino acid sequence and the backbone structure of the base can be summarized as follows.

(i-a) When the $(n-1)$th residue is not Asp, a kinked base is intrinsically formed due to the common $(n+1)$th Trp residue, which is able to form a hydrogen bond with the backbone carbonyl of the $(n-2)$th residue.

(i-b) When the $(n-1)$th residue is Asp and the 0th residue is neither Arg nor Lys, the carboxyl group of the $(n-1)$th Asp replaces the acceptor of the aforementioned hydrogen bond, and the extended structure is formed.

(i-c) When the $(n-1)$th residue is Asp, the 0th residue is Arg or Lys, and the (-1)th residue is neither Arg nor Lys, the side chain of the $(n-1)$th Asp is flipped to make a salt bridge with the basic side chain of the 0th residue, and forms the kinked base structure.

(i-d) When the $(n-1)$th residue is Asp, and both the 0th and the (-1)th residues are basic residues, the $(n-1)$th Asp forms a salt bridge with the (-1)th basic side chain, and makes the extended structure.

These rules from (i-a) to (i-d) agree in 52 of the 55 CDR-H3 structures, as indicated in Fig. 1. There are three exceptions. Two of them are the 26-10 antigen-free and complexed forms (PDB codes; 1IGI and 1IGJ), which were judged to have extended bases from rule (i-b), but the actual CDR-H3s form kinked base structures. In each structure, the side chain of the $(n-1)$th Asp makes a salt bridge with the $(n-5)$th Lys, thereby compensating for the lack of the 0th basic residue, and the characteristic hydrogen bond is formed between the backbone carbonyl of the $(n-2)$th residue and the side chain of the $(n+1)$th Trp. It is difficult to predict this salt bridge from only the amino acid sequence, but the concept of the rules may be correct in these cases. If the $(n-5)$th residue is not Lys, the loop might be extended without the flipping of the $(n-1)$th Asp.

The other exception is YST9.1 (1MAM), which was judged to have a kinked base from the above rule (i-a), but an extended base structure is observed. In this short CDR-H3 segment, composed of 8 residues, there are two Pro residues, which greatly deform the hairpin structure, as mentioned below. Probably, the effect of the Pro may be stronger than the effects of the base formation for relatively short CDR-H3 segments.

4. Insertion of an additional bulge above the base

Among the kinked base structures, additional bulges are sometimes inserted around the $(n-3)$th to the $(n-1)$th residues, just above the base. In fact, the scheme of the hydrogen bonds in the β-hairpin in SE155-4 (1MFA), BV04-01 (1NBV and 1CBV), and R45-45-11 (1IKF) changes from that in Fig. 3a to Fig. 3b, because of the inserted bulge. Then, new hydro-

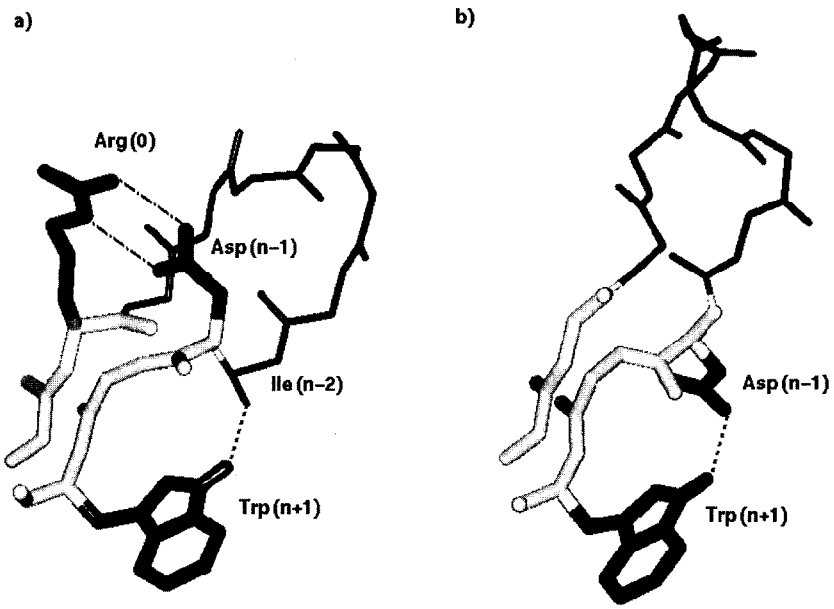

Fig. 2. Two different base structures in the CDR-H3 segments. The kinked base in NEW (a) and the extended base in 8F5 (b), with the same segment lengths ($n=9$). The thick white lines are the backbones of the bases, and the thick black lines are the side chains of the 0th Arg, the ($n-2$)th Asp, and the ($n+1$)th Trp. The thin lines are the β-hairpin backbones. The hydrogen bond between the HNε1 atom of the ($n+1$)th Trp and the ($n-2$)th carbonyl in (a), and that between the ($n+1$)th Trp and the ($n-1$)th Asp in (b), are indicated by dotted lines. A salt bridge between the 0th Arg and the (n−1)th Asp in (a) is indicated by a dot-dashed line.

gen bonds are able to form between the backbone amide of the 2nd residue and the backbone carbonyl of the ($n-4$)th residue, and between the carbonyl of the 2nd residue and the amide of the ($n-4$)th residue, instead of the ($n-3$)th residue.

In SE155-4, the ($n-2$)th residue is Gly. In most of the kinked base structures, a large hydrophobic residue, such as Phe or Met, is located at the ($n-2$)th position, and its side chain forms the bottom of the hydrophobic pocket of the antigen binding site [4]. Alternatively, in SE155-4, the side chain of the ($n-3$)th Tyr stays in the hydrophobic pocket, and compensates for the lack of the hydrophobic residue at the ($n-2$)th position.

In both BV04-01 and R45-45-11, the ($n-2$)th residue is Phe, but a Trp residue is at the ($n-3$)th residue. No space would be allowed for this bulky Trp side chain, due to the steric constraint with the ($n-1$)th backbone residue and the light chain, if it were located at the ($n-3$)th position in the normal, kinked structure shown in Fig. 3a.

Consequently, although the examples are few, the following relationship exists for the insertion of a bulge forming the modified kinked base, as shown in Fig. 3b. It is essential for classifying the β-hairpin conformations in the following section.

(ii) When either the ($n-3$)th residue is Trp, or the ($n-2$)th residue is Gly and the ($n-3$)th residue is a large hydrophobic residue, an additional bulge can be inserted in the kinked base.

5. Characteristics of the β-hairpin conformation

The CDR-H3 segments often show typical β-hairpins with significant hydrogen bond ladders between the adjacent backbone strands, which form a portion of the antiparallel sheet. However, several CDR-H3 lack hydrogen bonds, which results in largely deformed β-hairpin structures. It is one of the reasons why CDR-H3 segments are not considered to have any canonical structures. In fact, even the same antibody changes its CDR-H3 structure upon antigen binding [27–32], and it might be impossible to extract any relationship between the sequences and the β-hairpin conformations.

However, once the residue positions are assigned to one of the three β-hairpin schemes shown in Fig. 3a–c, several meaningful relationships appear. Now, let us introduce a new number, m, which equals $n-2$ for the kinked base without an additional bulge (Fig. 3a), $n-3$ for the kinked base with a bulge (Fig. 3b), and $n-1$ for the extended base (Fig. 3c). Then, the β-hairpins composed of m residues can be further grouped into the four basic classes from A to D, depending upon the value of m, in the conventional manner for globular proteins [35,36], as shown in Fig. 4. The relationships between the amino acid residues and these β-hairpin conformations were further studied. For very short CDR-H3 segments with $m \leq 4$, only a few examples were observed, and no remarkable relationships were found.

First, the position of the Pro residue and the β-hairpin conformation are closely related. In well-known conventional statistics [37], Pro tends to break an extended β-structure because of the lack of an amide group. In fact, the CDR-H3 segments in YST9.1 (1MAM), HIL (8FAB), and OPG2 (1OPG) include Pro residues, and the β-hairpins are largely deformed, and lack typical hydrogen bond ladders. In contrast, the CDR-H3 segments in KOL (2FB4) and R45-45-11(1IKF) form hydrogen bond ladders, which are accompanied by Pro.

These discrepancies originate in the position of Pro, as pointed out by Efimov [38]. The backbone structure of Pro is allowed in the extended structure. When the backbone N atom of Pro is not directed toward the opposite strand of its own β-hairpin, it is not necessary to break the β-hairpin with the typical hydrogen bond ladders. In the β-hairpin scheme

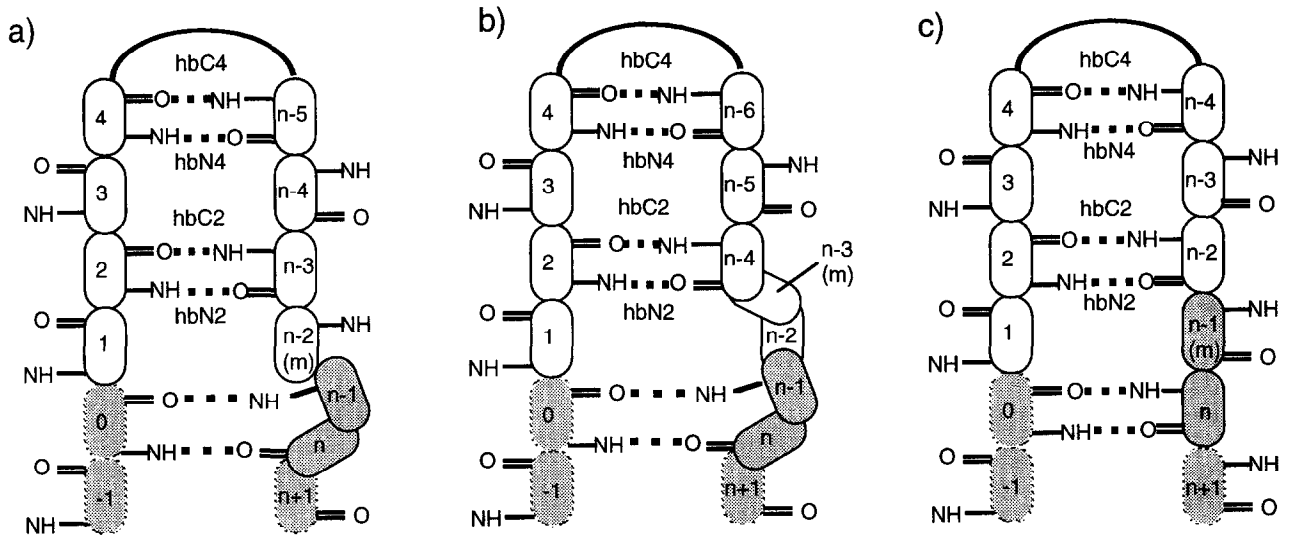

Fig. 3. Schematic representations of the CDR-H3 segments. The β-hairpin with a kinked base (a), with an additional bulge inserted in the kinked base (b), and the extended base structure (c). Ellipses with numbers from 1 to *n* are residues in the CDR-H3 segments defined by Kabat et al. [3]. Ellipses surrounded by dots are frame residues. Open ellipses belong to the β-hairpin region, and hatched ones belong to the base region. Hydrogen bonds constructing the ladder are indicated by broken lines. hbN*i* and hbC*i* (*i* = 2, 4, or 6) indicate hydrogen bonds, in which the N and C atoms of the *i*th residue participate, respectively.

shown in Fig. 3a–c, Pro at the 1st, the 3rd, the $(m-2)$th, or the mth position does not necessary break the hydrogen bond ladder. In contrast, when the backbone N atom of Pro is directed toward the opposite strand, it breaks the typical hydrogen bond ladder, without exception. The following rule is established.

(iii-a) When Pro is located at the specific residue positions 2, 4, m−3, or m−1, the β-hairpin structure is deformed and the hydrogen bond ladder should be broken.

Second, there is a correlation between the β-hairpin conformations with the kinked bases and the positions of the aromatic residues, which are abundant in the CDR segments [28]. Aromatic residues intrinsically prefer an extended structure to an α-helix [37]. The CDR-H3 segment with the kinked base lies in a narrow space surrounded by the other CDR segments and the frame structures. Therefore, when bulky aromatic residues are simultaneously located at adjacent positions on opposite strands, with the side chains extending in the same direction, the two antiparallel strands tend to be close together, resulting in a hydrogen bond ladder. This tendency is summarized as the following classification rule for the CDR-H3 segments with kinked bases:

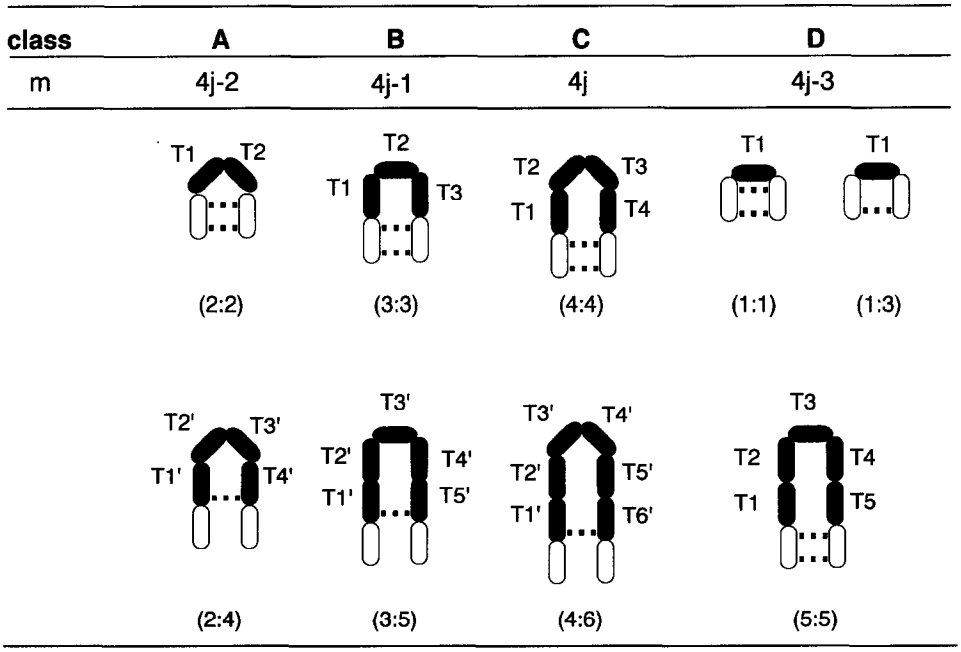

Fig. 4. Classification of β-hairpin structures depending on *m* with an integer *j*. The class names and the (X:Y) nomenclatures follow Sibanda et al. [36]. The filled ellipses are turn residues, and the broken lines indicate hydrogen bonds between the backbones.

(iii-b) When the base is kinked (Fig. 3a,b), and both the 2nd and the (m−1)th residues or both the 3rd and the (m−2)th residues are occupied by aromatic residues, a β-hairpin structure with a hydrogen bond ladder is constructed.

Third, detailed analyses of the turn conformations in the CDR-H3 segments were made, depending upon the group classification in Fig. 4. From the statistical investigations of β-hairpins in globular proteins [36], type I, type I′, and type II′ typical β-turns are frequently observed in class A. In these typical conformations, the loop residues T1 and T2 are often occupied by small amino acids, such as Gly, Asn, and Asp. In class B, the so-called 3:5 hairpin has a typical conformation, and the T4′ position is preferred only by Gly for the particular backbone structure of γ_L [36,39]. In class C, the 4:4 hair-

Table 2
Structural classification of the CDR-H3 segments and the relationships between the conformations and the amino acid sequences

Class[a]	Formation of (H-bond ladder, β-turn)[b]				
	m[c] (−,*)	(0,0)	(+,0)	(0,+)	(+,+)
(a) Kinked bases[d]					
A 6	**1MAM**(e; −,{7})[e]	**1PLG**(k; hbN2, I)		**1VFA**(k; hbC2, I′) **1VFB**(k; hbC2, I′) **1MFA**(k; hbC2, II)	
10	8FAB(k; −, I)	**1IGM**(k; hbN2,{6})			**6FAB**(k; hbC4, I′)
14	1OPG(k; −,{14})	**1IKF**(k; hbN6, I)			
B 3		**1IND**(k; −, II+L) **1GGC**(k; −,{3}) 1GGI(k; −, II+L)			
7		**1JEL**(k; −,{7}) **1NBV**(k; −,{7}) **1CBV**(k; hbC2,{3})		**7FAB**(k; hbN2, II′+L) **2FBJ**(k; hbN2, II+L) **1JHL**(k; hbN2,{3})	
11		**1LMK**(k; −,{4}) **1NMB**(k; hbN2,{4})	1IBG(k; hbN4,{3})		
15		**2FB4**(k; hbN2,{4})		**1DFB**(k; hbC2,{1})	
C 8		1IGF(k; hbN2,{4}) 2IGF(k; hbN2,{4}) 1FOR(k; −,II′) 1RMF(k; hbC2,{4}) **1DBA**(k; hbN2, I+R+L) **1DBJ**(k; hbN2, I+R+L)			
12			**1GIG**(k; hbC2,{8})		
D 5		**1CGS**(k; hbN2,{1}) **2CGR**(k; hbN2,{1}) **1TET**(k; hbN2,{1}) **1MLB**(k; hbC2,{1}) **1MLC**(k; hbC2,{1})			
9		1FVC(k; −,II) 1FVD(k; hbN2,{4}) **1FVE**(k; hbN2,{4}) **1NCA**(k; hbN2, II) **1HIL**(k; hbC2, III′) **1IFH**(k; −;{3}) **1FRG**(k; −,{3}) 1IGC(k; −;{2})	**1MCP**(k; hbN4,{1})		
13		1FGV(k; hbN2,{9})	1FAI(k; hbN2,{9})		
(b) Extended bases[d,f]					
A 6				1MRC(e; hbC2, II′) 1MRD(e; hbC2, II′) 1FLR(e; hbC2, I) **1BQL**(e; hbN2, I)	
C 8				1BBD(e; hbC2, I+R+L)	
D 9		1EAP(e; −, II′) **1IGI**(k; −, II′) 1IGJ(k; −,{8})			

[a]β-hairpin classes defined by Sibanda et al. [36].
[b]The positive and negative propensities of the hydrogen bond ladder and the β-turn formations are indicated by + and −, respectively, by rules from (iii) to (iv). 0 is neutral. (−,*) means that the hydrogen bond ladder is predicted to be broken regardless of β-turns from rule (iii-a).
[c]m equals $n-2$ for the kinked base without an additional bulge (Fig. 3a), $n-3$ for the kinked base with a bulge (Fig. 3b), and $n-1$ for the extended base (Fig. 3c).
[d]Kinked and extended bases were classified from only the amino acid sequences by rules from (i-a) to (i-d).
[e]The actual 3D structure of each CDR-H3 is indicated with the PDB code. The bold letter code corresponds to the well determined structure as that in Table 1. The base structures of the antibodies with underlines were erroneously predicted. The first item in the parentheses shows the kinked (k) or the extended (e) base structure. The second item is the ladder formation, indicating the farthest hydrogen bond from the base as hbNi or hbCi (i = 2, 4, or 6) defined in Fig. 3. The symbol '−' means no hydrogen bonds. The third item is the β-turn name. I, I′, II, II′, and III′ are type-I, type-I′, type-II, type-II′, and type-III′ turns, respectively. R and L are the right- and left-handed local helical conformations of the residue. {j} indicates that j residues form a loop, which is different from typical β-turn conformations.
[f]In Table 1, there was no sequence that was classified in (−,*) with an extended base. Since rule (iii-b) is only for kinked bases, neither (+,0) nor (+,+) groups appear in extended bases.

pin is frequently found, with the typical type I turn plus α_L structure at the T4 position, which is also preferred by Gly [36]. These tendencies are also observed in the CDR-H3 segments, when they form β-hairpins with remarkable hydrogen bond ladders. Thus, we add the final relationship.

(iv) When Gly, Asp, or Asn occupies in the position at T1 and/or T2 in class A, at T4' in class B, or at T4 in class C, a typical β-turn is formed.

6. Application of rules to structural modeling

Following the aforementioned rules, all of the CDR-H3 segments in Table 1 were classified from the sequences, and were compared with the actual 3D structures (Table 2), which are indicated by symbols in the parentheses after the PDB codes. As discussed above, the base structures were erroneously predicted for only three segments. For all of the other segments, the backbone structures of the bases were well determined from only the sequences. Since rule (iii-a) is strongly accompanied by the physical background, a CDR-H3 sequence that satisfies rule (iii-a) always has a deformed β-hairpin structure, without a significant hydrogen bond ladder. In addition, rules (iii-b) and (iv), as statistical tendencies, are compatible with formation of the β-hairpins. The β-hairpin conformations of 32 CDR-H3 segments with kinked bases and 3 segments with extended bases were left as ambiguous, because no remarkable features were observed in the amino acid sequences. In most cases, their β-hairpins actually lack long hydrogen bond ladders.

The amino acid sequences of CDR-H3 from the sequence database [3] were analyzed using the current rules, and the results are shown in Table 3. The extended bases were predicted to be distributed in every segment length, in contrast to those in the 3D structures, where only short segments ($n \leq 10$) were observed to have extended bases.

The ratio of the extended bases to all of the bases was 10.0%, and it is slightly smaller than the ratio of 13.6%, which was observed in the 3D structures shown in Table 2. In 8.0% of the predicted kinked bases, additional bulges were assumed to be inserted. The rate is similar to that in the 3D structures, 7.9%. These similarities suggest that at least the base structures are safely predicted by the current rules.

Although there still remain many CDR-H3 segments with ambiguous β-hairpin conformations, some segments have significant indications of typical β-hairpins, with hydrogen bond ladders with β-turns. The current relationship (iii-a) was observed from the 3D structures only in class A, but it could also discriminate the deformed β-hairpins from the sequence data in all classes from A to D; 14% of all kinked bases and 20% of all extended bases.

Table 3
Structural prediction for the CDR-H3 sequences [3]

Class[a]	m[c]	Formation of (H-bond ladder, β-turn)[b] (−, *)	(0,0)	(+,0)	(0,+)	(+,+)
(a) Kinked bases[d]						
A	6	25 (1)[e]	69 (6)	2	101 (17)	4
	10	39 (3)	71 (3)	14	79 (7)	55
	14	19 (2)	23 (2)	2	26	4
B	3	1	60	0	0	0
	7	23 (1)	156 (17)	14	60 (12)	7
	11	39 (5)	96 (14)	22 (2)	28 (1)	4
	15	13	23	4	8	2
C	4	2	45	0	5	0
	8	25 (1)	165 (17)	35 (1)	57 (8)	2
	12	30	71 (2)	11 (2)	12 (2)	4
D	5	12	115 (12)	7	0	0
	9	36 (4)	181 (14)	36	0	0
	13	24 (1)	57 (4)	7 (2)	6	0
Sum		288 (18)	1132 (91)	154 (7)	382 (47)	82 (0)
Total kinked bases: 2038 (163)						
(b) Extended bases[d,f]						
A	6	2	8		14	
	10	1	10		6	
	14	9	7		2	
B	7	1	23		5	
	11	5	15		1	
	15	3	4		0	
C	4	0	0		2	
	8	2	21		5	
	12	2	4		3	
	16	5	4		1	
D	5	2	11		0	
	9	8	28		0	
	13	5	6		0	
Sum		45	141		39	
Total extended bases: 225						

The numbers of the CDR-H3 segments predicted to be classified by their sequence features.
[a]β-hairpin classes defined by Sibanda et al. [36].
[b]The positive and negative propensities of the hydrogen bond ladder and the β-turn formations are indicated by + and -, respectively, by rules from (iii) to (iv). 0 is neutral. (−,*) means that the hydrogen bond ladder is predicted to be broken regardless of β-turns from rule (iii-a).
[c]m equals $n-2$ for the kinked base without an additional bulge (Fig. 3a), $n-3$ for the kinked base with a bulge (Fig. 3b), and $n-1$ for the extended base (Fig. 3c).
[d]Kinked and extended bases were classified from only the amino acid sequences by rules from (i-a) to (i-d).
[e]Numbers are results for total 2263 sequences from the Kabat Database [3] with lengths from 5 to 17, where the identical CDR-H3 sequences were removed. Numbers in parentheses are the predicted cases with the insertion of an additional bulge from rule (ii).
[f]Since rule (iii-b) is only for kinked bases, neither (+,0) nor (+,+) groups appear in extended bases.

7. Structural diversity of CDR-H3

Structural changes in CDR-H3 often accompany antigen binding [27–32]. In the current 3D structures, the coordinates of 10 antibodies are available for both the antigen-free and the complexed structures. A comparative study revealed that seven of the antibodies (JEL103, NC6.8, D44.1, D1.3, B13I2, DB3, and 26-10) do not display any significant changes in their backbone conformations upon antigen binding, and their backbone root mean square deviations (r.m.s.d.s) of the CDR-H3 are less than 0.8 Å. In contrast, significant structural differences between the free and the complexed forms have been observed in three antibodies (BV04-01 [30], 17/9 [31], and 50.1 [32]), although the kinked base structures are always maintained. The CDR-H3 segment of BV04-01 has a typical β-hairpin in the antigen-complexed form (PDB code; 1CBV), but it lacks the hydrogen bond ladder in the antigen-free form (1NBV). The overall and base backbone r.m.s.d.s of the CDR-H3 segment are 1.2 and 0.4 Å, respectively [30]. Both segment structures of 17/9 and 50.1 in the antigen-free and complexed forms lack typical hydrogen bond ladders in their β-hairpins, and they show large changes in the segment con-

formations, with backbone r.m.s.d.s of 1.9 and 1.5 Å, respectively [31,32].

The structural comparison between the CDR-H3 segments, whose amino acid sequences are very similar to each other, also provides information about the structural diversity. 4D5 ver 7 (1FVE) has the same CDR-H3 amino acid sequence as that of 4D5 ver 8 (1FVC), and one residue at the nth position in 4D5 ver 4 (1FVD) is substituted in 4D5 ver 8. Although the kinked base structures are similar in all three antibodies, with a base backbone r.m.s.d. ≤ 0.4 Å, the overall CDR-H3 segments show very different backbone conformations. In addition, only one residue differs in the sequence of 26/9 from that of 17/9, and the entire complexed CDR-H3 structure of 26/9 (1FRG) is similar to the complexed structure of 17/9 (1IFH), with only a 0.2 Å backbone r.m.s.d., but it is very different from the free form of 17/9 (1HIL), with a 1.8 Å backbone r.m.s.d.

It should be emphasized that all of this structural diversity occurs only for the CDR-H3 segments, whose amino acid sequences in the β-hairpins lack features as remarkable as the current rules for the β-hairpins, as shown in Table 2. On the contrary, little structural diversity is expected for the CDR-H3 segments which have significant indications of typical β-hairpins, with hydrogen bond ladders and β-turns.

8. Conclusion

We classified the 3D local structures of the CDR-H3 segments of many antibodies, and found several significant relationships between the amino acid sequences and the loop conformations. Especially the discrimination between the kinked and extended bases from only the sequences is almost completely compatible with the actual conformations. Several β-hairpin structures correlate well with the sequences, when remarkable indications are observed for the hydrogen bond ladders and the typical β-turns. However, other β-hairpins should have flexible conformations, suitable for antigen recognition. For these segments, instead of trying to find further structural rules, powerful conformation search calculations [40] are expected to provide putative structural models for the β-hairpins, constructed on the rigid kinked or extended base structures.

Acknowledgements: The authors thank Dr. M. Suyama (BERI) for his assistance for Kabat Database search.

References

[1] Davies, D.R. and Padlan, E.A. (1990) Annu. Rev. Biochem. 59, 439–473.
[2] Padlan, E.A. (1994) Mol. Immunol. 31, 169–217.
[3] Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S. and Foeller, C. (1991) Sequences of Proteins of Immunological Interest, 5th edn. U.S. Department of Health and Human services, Public Health Service, National Institutes of Health, NIH Publication No. 91–3242, Bethesda, MD.
[4] Chothia, C. and Lesk, A.M. (1987) J. Mol. Biol. 196, 901–917.
[5] Chothia, C., Lesk, A.M., Tramontano, A., Levitt, M., Smith-Gill, S.J., Air, G., Sheriff, S., Padlan, E.A., Davies, D., Tulip, W.R., Colman, P.M., Spinelli, S., Alzari, P.M. and Poljak, R.J. (1989) Nature 342, 877–883.
[6] Tramontano, A., Chothia, C. and Lesk, A.M. (1989) Proteins Struct. Funct. Genet. 6, 382–394.
[7] Chothia, C., Lesk, A.M., Gherardi, E., Tomlinson, I.M., Walter, G., Marks, J.D., Llewelyn, M.B. and Winter, G. (1992) J. Mol. Biol. 227, 799–817.
[8] Wu, S. and Cygler, M. (1993) J. Mol. Biol. 229, 597–601.
[9] Vargas-Madrazo, E., Lara-Ochoa, F. and Almagro, J.C. (1995) J. Mol. Biol. 254, 497–504.
[10] Vargas-Madrazo, E., Lara-Ochoa, F. and Almagro, J.C. (1996) J. Mol. Biol. 258, 893.
[11] Höhne, W.E., Küttner, G., Kießig, S., Hausdorf, G., Grunow, R., Winkler, K., Wessner, H., Gießmann, E., Stigler, R., Schneider-Mergener, J., von Baehr, R. and Schomburg, D. (1993) Mol. Immunol. 30, 1213–1221.
[12] Roberts, V.A., Stewart, J., Benkovic, S.J. and Getzoff, E.D. (1994) J. Mol. Biol. 235, 1098–1116.
[13] Barry, M.M., Mol, C.D., Anderson, W.F. and Lee, J.S. (1994) J. Biol. Chem. 269, 3623–3632.
[14] Reczko, M., Martin, A.C.R., Bohr, H. and Suhai, S. (1995) Protein Eng. 8, 389–395.
[15] Mandal, C., Kingery, B.D., Anchin, J.M., Subramaniam, S. and Linthicum, D.S. (1996) Nature Biotech. 14, 323–328.
[16] Bruccoleri, R.E., Haber, E. and Novotny, J. (1988) Nature 335, 564–568.
[17] Higo, J., Collura, V. and Garnier, J. (1992) Biopolymers 32, 33–43.
[18] Zheng, Q., Rosenfeld, R., Delisi, C. and Kyle, D.J. (1994) Protein Sci. 3, 493–506.
[19] Martin, A.C.R., Cheetham, J.C. and Rees, A.R. (1989) Proc. Natl. Acad. Sci. USA. 86, 9268–9272.
[20] Webster, D.M. and Rees, A.R. (1995) Methods Mol. Biol. 51, 17–49.
[21] Mas, M.T., Smith, K.C., Yarmush, D.L., Aisaka, K. and Fine, R.M. (1992) Protein Struct. Funct. Genet. 14, 483–498.
[22] Fujii, I., Tanaka, F., Miyashita, H., Tanimura, R. and Kinoshita, K. (1995) J. Am. Chem. Soc. 117, 6199–6209.
[23] Bajorath, J. and Sheriff, S. (1996) Proteins Struct. Funct. Genet. 24, 152–157.
[24] Wu, T.T., Johnson, G. and Kabat, E.A. (1993) Proteins Struct. Funct. Genet. 16, 1–7.
[25] Segal, D.M., Padlan, E.A., Cohen, G.H., Rudikoff, S., Potter, M. and Davies, D.R. (1974) Proc. Natl. Acad. Sci. USA 71, 4298–4302.
[26] Sharon, J. (1990) Proc. Natl. Acad. Sci. USA 87, 4814–4817.
[27] Wilson, I.A. and Stanfield, R.L. (1994) Curr. Opin. Struct. Biol. 4, 857 867.
[28] Mian, I.S., Bradwell, A.R. and Olson, A.J. (1991) J. Mol. Biol. 217, 133–151.
[29] Odaka, A., Kim, J.I., Takahashi, H., Shimada, I. and Arata, Y. (1992) Biochemistry 31, 10686–10691.
[30] Herron, J.N., He, X.M., Ballard, D.W., Blier, P.R., Pace, P.E., Bothwell, A.L.M., Voss, E.W., Jr. and Edmundson, A.B. (1991) Proteins Struct. Funct. Genet. 11, 159–175.
[31] Rini, J.M., Schulze-Gahmen, U. and Wilson, I.A. (1992) Science 255, 959–965.
[32] Stanfield, R.L., Takimoto-Kamimura, M., Rini, J.M., Profy, A.T. and Wilson, I.A. (1993) Structure 1, 83–93.
[33] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[34] Chien, N.C., Roberts, V.A., Giusti, A.M., Scharff, M.D. and Getzoff, E.D. (1989) Proc. Natl. Acad. Sci. USA 86, 5532–5536.
[35] Milner-White, E.J. and Poet, R. (1986) Biochem. J. 240, 289–292.
[36] Sibanda, B.L., Blundell, T.L. and Thornton, J.M. (1989) J. Mol. Biol. 206, 759–777.
[37] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. 47, 45–148.
[38] Efimov, A.V. (1991) FEBS Lett. 284, 288–292.
[39] Ishikawa, K., Kimura, S., Kanaya, S., Morikawa, K. and Nakamura, H. (1993) Protein. Eng. 6, 85–91.
[40] Kidera, A. (1995) Proc. Natl. Acad. Sci. USA 92, 9886–9889.