FEBS Letters 455 (1999) 188–197 FEBS 22273

H3-rules: identification of CDR-H3 structures in antibodies

Hiroki Shirai[1], Akinori Kidera[2], Haruki Nakamura*

Department of Bioinformatics, Biomolecular Engineering Research Institute, 6-2-3 Furuedai, Suita, Osaka 565-0874, Japan

Received 31 March 1999; received in revised form 9 June 1999

Abstract For the third complementarity determining region of the antibody heavy chain (CDR-H3), we propose the 'H3-rules', which should identify the tertiary structure from the amino acid sequence of the CDR-H3 segment. A total of 100 CDR-H3 segments from well-determined crystal structures were analyzed. Distinctive relationships between the structures and the sequences were revealed from 55 segments, and the rules were examined for the other 45 segments and were verified. In some antibodies, basic residues at specific positions were revealed to be notable signals, with their ability to form salt bridges and to assume conformations inconsistent with the rules.

Key words: Loop modeling; Antibody structure; Antibody engineering; Combining site

1. Introduction

The antigen binding site of an antibody is composed of six complementarity determining regions (CDRs), namely L1, L2, L3, H1, H2, and H3, with a large structural repertoire [1–3], and its capacity for tight and specific antigen binding has been widely applied in therapeutics [4]. Among the six CDR segments, the H3 segment is produced by joining and editing of the V-D-J sequences, and it has the largest variety in its length, sequence, and structure [5,6]. The CDR-H3 segment has a distinctive role in antigen recognition, and sometimes changes its conformation upon antigen binding [7].

In order to understand the mechanism of immune response based upon the tertiary structure of an antibody, the architecture of those CDR segments is crucial. Five of the CDR segments, except CDR-H3, were classified into a small number of 'canonical structures' by their amino acid sequences [5,6]. In contrast, no such 'canonical structures' have ever been established for the CDR-H3 fragments [6,7], due to the large conformational variety. Identification of the CDR-H3 structure is therefore an essential issue for the analysis of immune response and for antibody engineering.

We had analyzed the 55 well-determined CDR-H3 crystal structures by careful inspection, and had derived several rules as a 'hypothesis', indicated in the upper half of Table 1, which could govern the CDR-H3 conformation [8]. After this proposal, other attempts were also made to reveal the relationships between the sequences and the structures of the CDR-H3. Morea et al. performed an extensive database analysis by careful inspection [9], which resulted in essentially the same conclusions as ours, probably because the same structural data were investigated. Oliva et al. analyzed the CDR-H3 structures by an automatic clustering procedure [10].

The number of antibody crystal structures is rapidly increasing. Here, 45 additional antibody structures were analyzed as a blind test of the previous rules, and remarkable signals informing us of putative exceptional cases were revealed.

2. Materials and methods

The atomic coordinates of the antibodies were obtained from the Brookhaven Protein Data Bank (PDB) [11]. The structures of the highest resolutions were first selected as representatives for the free and the complexed antibodies. Then, a total of 45 CDR-H3 segment structures with resolutions equal to or better than 2.8 Å was prepared from the new entries from releases #77 to #85 plus others registered before October, 1998, for these representatives (Table 2). The individual H3-rules were examined for the 100 CDR-H3 segments. The structures were displayed on a computer graphics system (Indy-XZ; Silicon Graphics Inc., CA, USA), and were observed with the graphics program Insight-II (Molecular Simulations Inc., San Diego, CA, USA).

3. Results and discussion

3.1. Original rules for the CDR-H3 structures

The relationships between the sequences and the structures were previously proposed [8] as indicated in the upper half of Table 1, and they are described briefly. The length n of the CDR-H3 segment is so variable that the segment residues are re-numbered from 1 to n, corresponding to the conventional residue numbers [1] from 95 to 102 of the heavy chain. The CDR-H3 structures are divided into two regions, 'base' and 'β-hairpin', which are proximal and distal to the framework, respectively (Fig. 1).

The bases could be classified into two forms, a kinked base K that contains a β-bulge at the $(n-1)$st residue, and an extended base E that lacks a β-bulge but forms normal antiparallel β-strands (Figs. 1 and 2a). A salt bridge between the side chain of the $(n-1)$st Asp and the N-terminal basic residue could switch the two base forms, and we proposed the rules from i-a to i-d, as indicated in Table 1. From rule ii, the conformations with additional bulges inserted just above the bases can be identified, and are named K^+ [10]. From the number of residues m in the β-hairpin region, the β-hairpin classes A to D [12] are identified (Fig. 2b). By examining the compatibility of the sequence with the β-hairpin structures, we

*Corresponding author. Present address: Institute for Protein Research, Osaka University, Yamadaoka, Suita, Osaka 565-0871, Japan. Fax: (81) (6) 6879-4310.
E-mail: harukin@protein.osaka-u.ac.jp

[1] Present address: Tanabe Seiyaku Co., Ltd., Kashima, Yodogawa-ku, Osaka 532-8505, Japan.
[2] Present address: Graduate School of Science, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606-8502, Japan.

Abbreviations: CDR, complementarity determining region; H3, complementarity determining region 3 from heavy chain; rmsd, root-mean-square deviation

PII: S0014-5793(99)00821-2

Table 1
H3-rules

Original rules	
i.	Base type identification
i-a	When position ($n-1$) is not Asp, a kinked base (K) is formed.
i-b	When position ($n-1$) is Asp and (0) is not a basic residue, an extended base (E) is formed.
i-c	When position ($n-1$) is Asp, (-1) is not basic, and the (0) residue is basic, a kinked base is formed.
i-d	When position ($n-1$) is Asp and both (-1) and (0) residues are basic, an extended base is formed.
ii.	Further classification of each base type Either when position ($n-3$) is Trp, or when position ($n-2$) is Gly and ($n-3$) is a large hydrophobic residue, the second bulge is additionally inserted in the kinked base.
iii.	Formation of the hydrogen bond ladder
iii-a	When Pro is located at positions 2, 4, $m-3$, or $m-1$, the β-hairpin should be broken.
iii-b	When the base is kinked, the above rule iii-a is not satisfied, and both the 2nd and the ($m-1$)st residues or both the 3rd and ($m-2$)nd residues are occupied by aromatic residues, a β-hairpin with a hydrogen bond ladder is formed.
iv.	Formation of the β-turn When the above rule iii-a is not satisfied, and Gly, Asp, or Asn is located at the position of (m/2 or m/2+1 in class A),at ((m+3)/2 in class B), or at (m/2+2 in class C), a typical β-turn in the corresponding class is formed.
Additional rules	
Notable signals for rule i-b	When the sequence satisfies the condition of rule (i-b), if one of the residues between positions 1 and ($n-3$) is a basic residue, or L49 is a basic residue, then the side chain of the ($n-1$)st Asp often makes a salt bridge with these basic residues, and a kinked base is formed.
Notable signals for rule i-c	When the sequence satisfies the condition of rule ($i-c$), if L46 is a basic residue, then the ($n-1$)st Asp often makes a salt bridge with the L46 basic side chain, and an extended base is formed.
ii-a	If the base is kinked, when position ($n-3$) is Gly, then the K^G type of stem is formed. The pseudo-dihedral angle Θ_{stem} is larger than $-100°$ and smaller than $-10°$, assuming the (φ, ψ) backbone angle of the Gly is in either the irregular region, ε, or a very extended conformation. (Θ_{stem} is the pseudo-dihedral angle formed by the four successive C^α atoms from the ($n-4$)th to the ($n-1$)st residues.)
ii-b	If the base is kinked, either when position ($n-3$) is Trp, or when ($n-2$) is Gly and ($n-3$) is a large hydrophobic residue, then a K^+-type stem is formed. The Θ_{stem} is from $-10°$ to $50°$.
ii-c	If the base is extended, when position (0) is Gly or Arg, then an extended base of the class $E^N$with a Θ_{base} value around -160 ° is formed. In contrast, when position (0) is neither Gly nor Arg, the class E^P with the Θ_{base} around 160 ° is formed. (Θ_{base} is the pseudo-dihedral angle formed by the four successive C^α atoms from the ($n-2$)nd to the ($n+1$)st residues.)

Table 2
100 CDR-H3 structures

Antibody name	PDB code (+/-)[1]	Sequence[2] CDR-H3	L46	L49	n[3]	m[4]	Predicted structures / Crystal structures[5] rule i[6]	rule ii[7]	rule iii[8]	rule iv[9]	Θ_{base}[10]	Θ_{stem}[11]	r.ms.d.[12] A(Å)	B(Å)
Data set 1[13]		(1 n)												
CHA255	1IND +	CAS----------(HRFVH)W	G	G	5	3	K/K	-	-	*/T:II+L	41	-		
50.1	1GGC -	CAQ----------(EGYIY)W	L	Y	5	3	K/K	-	-	*/T:II+L	20	-		
50.1	1GGI +	CAQ----------(EGYIY)W	L	Y	5	3	K/K	-	-	*/T:II+L	63	-	1.5	0.3
JEL103	1MRC -	CAN----------(LRGYFDY)W	L	Y	7	6,5	E,K/E	P,*/P	*,*/L:C2	T,*/T:II'	159	-		
JEL103	1MRD +	CAN----------(LRGYFDY)W	L	Y	7	6,5	E,K/E	P,*/P	*,*/L:C2	T,*/T:II'	159	-	0.4	
4-4-20	1FLR +	CTG----------(SYYGMDY)W	V	Y	7	6	E/E	N/N	*/L:C2	T/T:I	-162	-		
NC6.8	1CGS -	CTR----------(GYSSMDY)W	L	Y	7	5	K/K	*/G	*/L:N2	*/{1}	-0	-92		
NC6.8	2CGR +	CTR----------(GYSSMDY)W	L	Y	7	5	K/K	*/T	*/L:N2	*/{1}	35	-117	0.8	
HyHEL5	1BQL +	CLH----------(GNYDFDG)W	R	Y	7	6	E/E	P/P	*/L:N2	T/T:I	150	-		
TE33	1TET +	CAR----------(RSWYFDV)W	L	Y	7	5	K/K	*/G	*/L:N2	*/{1}	38	-89		
D44.1	1MLB -	CAR----------(GDGNYGY)W	L	K	7	5	K/K	*/T	*/L:C2	*/{1}	56	-123		
D44.1	1MLC +	CAR----------(GDGNYGY)W	L	K	7	5	K/K	*/T	*/L:C2	*/{1}	52	-122	0.4	0.2
D1.3	1VFA -	CAR----------(ERDYRLDY)W	L	Y	8	6	K/K	*/T	*/L:C2	T/T:I'	36	-131		
D1.3[14]	1A2Y +	CAR----------(ERDYRLDY)W	L	Y	8	6	K/K	*/T	*/L:C2	T/T:I'	29	-122	0.4	
YST9.1	1MAM -	CTR----------(DPYGPAAY)W	L	Y	8	7	K/E	*/P	D/D	*/{7}	163	-		
PLG	1PLG -	CAR----------(GGKFAMDY)W	P	Y	8	6	K/K	*/T	*/L:N2	*/T:I	36	-118		
D11.15	1JHL +	CAR----------(DDNYGAMDY)W	L	Y	9	7	K/K	*/T	*/L:N2	T/{3}	33	-119		
J539	2FBJ +	CAR----------(LHYYGYNAY)W	P	Y	9	7	K/K	*/T	*/L:N2	T/T:II+L	27	-135		
SE155-4	1MFA +	CTR----------(GGHGYYGDY)W	G	G	9	6	K/K	+/+	*/L:C2	T/T:II	31	22		
JEL142[15]	2JEL +	CAR----------(VMGEQYFDV)W	L	Y	9	6	K/K	*/+	*/D	*/{6}	34	18		
NEW	7FAB -	CAR----------(NLIAGGIDV)W	L	-	9	7	K/K	G/G	*/L:N2	T/T:II'+L	42	-47		
8F5	1BBD -	CDG----------(YYSYYDMDY)W	L	Y	9	8	E/E	N/N	*/L:C2	T/T:I+R+L	-152	-		
B13I2	1IGF -	CTR----------(YSSDPFYFDY)W	L	Y	10	8	K/K	*/T	*/L:N2	*/{4}	40	-114		0.4
B13I2	2IGF +	CTR----------(YSSDPFYFDY)W	L	Y	10	8	K/K	*/T	*/L:N2	*/{4}	30	-113	0.5	
17-IA	1FOR -	CAR----------(SGNYPYAMDY)W	L	Y	10	8	K/K	*/T	*/D	*/{8}	31	-125		
R6.5	1RMF -	CAR----------(GGWLLLSFDY)W	L	Y	10	7	K/K	*/C	*/L:C2	*/{4}	30	22		
DB3	1DBA -	CTR----------(GDYVNWYFDV)W	L	Y	10	8	K/K	*/T	L/L:N2	*/T:I+R+L	45	-118		
DB3	1DBJ +	CTR----------(GDYVNWYFDV)W	L	Y	10	8	K/K	*/T	L/L:N2	*/T:I+R+L	44	-114	0.4	
BV04-01	1NBV -	CVR----------(DQTGTAWFAY)W	L	Y	10	7	K/K	+/+	*/D	*/{7}	25	17		
BV04-01	1CBV +	CVR----------(DQTGTAWFAY)W	L	Y	10	7	K/K	+/+	*/T:C2	*/{3}	49	36	1.2	
17E8	1EAP +	CKR----------(SYYGSSYVDY)W	L	H	10	9	E/E	N/N	*/D	*/T:II'	-139	-		
26-10	1IGI -	CAG----------(SSGNKWAMDY)W	L	Y	10	9,8	E,K/K	N,*/T	*,*/D	*,*/T:II'	38	-109		

Table 2 (*continued*)

26-10	1IGJ	+	CAG---------(SSGNKWAMDY)W	L	Y	10	9,8	E,K/K	N,*/T	*,*/D	*,*/{8}	50	-120	0.6	0.4	
MOPC21	1IGC	-	CAR--------(WGNYPYYAMDY)W	L	Y	11	9	K/K	*/G	*/D	*/{2}	19	-79			
4D5ver4	1FVD	-	CSR--------(WGGDGFYAMDV)W	L	Y	11	9	K/K	*/G	*/L:N2	*/{4}	49	-17		2.7	
4D5ver7	1FVE	-	CSR--------(WGGDGFYAMDY)W	L	Y	11	8	K/K	*/C	*/L:N2	*/{4}	41	11		2.7	
4D5ver8	1FVC	-	CSR--------(WGGDGFYAMDY)W	L	Y	11	9	K/K	*/T	*/D	*/T:II	41	-111		0.6	
17/9	1HIL	-	CAR--------(RERYDENGFAY)W	V	Y	11	9	K/K	G/G	*/L:C2	*/T:III'	19	-68		0.1	
17/9	1IFH	+	CAR--------(RERYDENGFAY)W	V	Y	11	9	K/K	G/G	*/D	*/{3}	48	-25	1.9		
26/9	1FRG	+	CAR--------(RERYDEKGFAY)W	L	Y	11	9	K/K	G/G	*/D	*/{3}	55	-26			
McPC603	1MCP	-	CAR--------(NYYGSTWYFDV)W	L	Y	11	9	K/K	*/T	L/L:N4	*/{1}	30	-105			
NC41	1NCA	+	CAR--------(GEDNFGSLSDY)W	L	Y	11	9	K/K	*/T	*/D	*/{2}	24	-162			
36-71	6FAB	-	CAR-------(SEYYGGSYKFDY)W	L	Y	12	10	K/K	*/T	L/L:C4	T/T:I'	29	-130			
HIL	8FAB	-	CAR-------(DPDILTAFSFDY)W	M	Y	12	10	K/K	*/T	D/D	*/{10}	31	-122		1.6	
POT	1IGM	-	CAR-------(HRVSYVLTGFDS)W	L	Y	12	10	K/K	G/G	*/L:N2	*/{6}	16	-94			
L5MK161	1LMK	-	CAR------(GEDYYAYWYVLDY)W	L	Y	13	11	K/K	*/T	*/D	*/{4}	54	-128		0.4	
NC10	1NMB	+	CAR------(SGGSYRYDGGFDY)W	L	Y	13	11	K/K	G/G	*/L:N2	*/{4}	31	-49			
40-50	1IBG	+	CAR------(FRFASYYDYAVDY)W	L	Y	13	11	K/K	*/T	*/L:N4	*/{3}	40	-129			
HC19	1GIG	-	CAR-----(DFYDYDVFYYAMDY)W	G	G	14	12	K/K	*/T	L/L:C2	*/{8}	29	-132			
H52	1FGV	-	CAR----(WRGLNYGFDVRYFDV)W	L	Y	15	13	K/K	*/T	*/L:N2	*/{9}	35	-105			
R19.9	1FAI	-	CAR----(SFYGGSDLAVYYFDS)W	L	Y	15	13	K/K	*/G	L/L:N2	*/{9}	32	-89			
OPG2	1OPG	-	CTR---(HPFYRYDGGNYYAMDH)W	L	K	16	14	K/K	*/T	D/D	*/{14}	38	-127		2.0	
KOL	2FB4	-	CAR--(DGGHGFCSSASCFGPDY)W	L	Y	17	15	K/K	G/G	L/L:N2	*/{4}	28	-24			
3D6	1DFB	-	CVK--(GRDYYDSGGYFTVAFDI)W	L	Y	17	15	K/K	*/T	*/L:C2	T/{1}	48	-131			
R45-45-11	1IKF	+	CTR--(HTLYDTLYGNYPVWFAD)W	L	F	17	14	K/K	+/+	*/L:N6	*/T:I	25	46			
Data set 2[13)]																
48G7	1HKL	-	CAS--------------(YYGIY)W	R	Y	5	3	K/K	-	-	*/T:II+L	29	-			
48G7	1GAF	+	CTS--------------(YYGIY)W	R	Y	5	3	K/K	-	-	*/T:II+L	31	-	0.1		
M41	1GPO	-	CAN--------------(WHGDY)W	L	K	5	4,3	E,K/K	-	-	*,*/{3}	22	-		0.2	
ME36.1	1PSK	-	CTS--------------(KSFDY)W	L	Y	5	4,3	E,K/K	-	-	*,*/T:II+L	48	-			
H1002	1GHF	-	CAR------------(VEAGFDY)W	L	Y	7	5	K/K	G/G	L/L:C2	*/{1}	9	-45			
28B4	1KEL	-	CAR-----------(WGSYAMDY)W	L	Y	8	6	K/K	*/G	*/L:C2	*/T(I)	44	-12			
28B4	1KEM	+	CAR-----------(WGSYAMDY)W	L	Y	8	6	K/K	*/G	*/L:C2	*/T(I)	44	-14	0.3		
D1.3mut	1KIQ	+	CAR-----------(ERDFRLDY)W	L	Y	8	6	K/K	*/T	*/L:N2	*/T:I'	29	-121			
5G9	1FGN	-	CAR-----------(DNSYYFDY)W	T	Y	8	6	K/K	*/G	*/L:N2	*/{4}	34	-75			
RFT5	1MIM	-	CSR-----------(DYGYYFDF)W	R	Y	8	6	K,E/K	*,N/T	L,*/L:N2	T,*/T:II'	34	-113			
MAB1-IA	1A6T	-	CAR-----------(RDDYYFDF)W	P	Y	8	6	K/K	*/G	L/L:N2	T/T:I	32	-66		0.7	
7E2	1AR1	+	CVR----------(HEYYYAMDY)W	F	Y	9	7	K/K	*/T	L/L:C2	*/{3}	44	-129			

Table 2 (*continued*)

H57	1NFD	+	CTR----------(AGRFDHFDY)W	L	Y	9	7	K/K	*/G	L/L:N2	T/T:I	46	-25		0.3	
8F5	1A3R	+	CDG----------(YYSYYDMDY)W	L	Y	9	8	E/E	N/N	*/L:C2	T/T:I	-166	-	0.5		
MAB231	1IGT	-	CAR----------(HGGYYAMDY)W	L	Y	9	7	K/K	*/T	*/L:N2	*/{3}	34	-106		0.6	
BR96	1CLY	+	CAR---------(GLDDGAWFAY)W	L	Y	10	7	K/K	+/+	*/L:C2	T/T:I+L	40	20	0.4		
BR96	1UCB	-	CAR---------(GLDDGAWFAY)W	L	Y	10	7	K/K	+/+	*/L:C2	T/T:I+L	37	23			
DESIRE-1	1KB5	+	CAR---------(SRTDLYYFDI)W	L	Y	10	8	K/K	*/T	*/D	*/{8}	10	-114			
FV4155	1CFV	+	CAR---------(LNYAVYGMDY)W	L	Y	10	8	K/K	G/G	L/L:C2	*/T:I+R+L	29	-37			
A5B7	1CLO	-	CTR---------(DRGLRFYFDY)W	S	Y	10	8	K/K	*/T	*/L:N2	*/T:II'	32	-106			
39-A11	1A4J	-	CVQ---------(AERLRRTFDY)W	L	Y	10	9,8	E,K/K	P,*/T	*,*/L:C2	*,*/T:I+R+L	26	-117		0.2	
39-A11	1A4K	+	CVQ---------(AERLRRTFDY)W	L	Y	10	9,8	E,K/K	P,*/T	*,*/L:C2	*,*/T:I+R+L	22	-127	0.2	0.1	
TP7	1AY1	-	CAR---------(YYYGYWYFDV)W	L	Y	10	8	K/K	*/T	L/L:N2	*/{1}	45	-105			
N1G9	1NGP	+	CAR--------(YDYYGSSYFDY)W	G	G	11	9	K/K	*/T	*/L:N2	*/{5}	33	-126	0.2		
N1G9	1NGQ	-	CAR--------(YDYYGSSYFDY)W	G	G	11	9	K/K	*/T	*/L:N2	*/{5}	40	-123			
B1-8	1A6U	-	CAR--------(YDYYGSSYFDY)W	G	G	11	9	K/K	*/T	*/L:N2	*/{5}	38	-132			
B1-8	1A6W	+	CAR--------(YDYYGSSYFDY)W	G	G	11	9	K/K	*/T	*/L:N2	*/{5}	37	-124	0.3		
CTM01	1AD9	-	CAR--------(EKTTYYYAMDY)W	L	Y	11	9	K/K	*/T	*/L:C2	*/{5}	52	-150		0.8	
59.1	1AI1	+	CSR--------(ENHMYETYFDV)W	V	Y	11	9	K/K	*/T	*/D	*/{9}	39	-105			
2E8	12E8	-	CNA--------(GHDYDRGRFPY)W	L	Y	11	9	K/K	*/T	*/D	*/{9}	49	-149		0.4	
C219	1AP2	-	CAR-------(REVYSYYSPLDV)W	L	Y	12	10	K/K	*/G	D/D	*/{10}	52	-50		0.4	
A6	1JRH	-	CAR-------(RAPFYGNHAMDY)W	L	S	12	10	K/K	*/T	*/L:N2	T/{6}	41	-112			
184.1	1OSP	+	CAR-------(SRDYYGSSGFAF)W	L	S	12	10	K/K	G/G	*/L:N2	T/{6}	43	-91			
MN12H2	1MPA	+	CSI-------(IYFDYADFIMDY)W	L	Y	12	11	E/E	P/P	*/L:C2	*/{7}	151	-			
TR1.9	1VGE	-	CAR-------(DPYGGGKSEFDY)W	L	Y	12	10	K/K	*/T	D/D	*/{10}	36	-116			
B1	1DSF	-	CGR-------(SPIYYDYAPFTT)W	L	Y	12	9	K/K	*/+	D/D	*/{9}	27	28			
E5.2	1DVF	+	CAT------(KVIYYQGRGAMDY)W	L	Y	13	12,11	E,K/K	P,*/T	*,*/D	*,*/{11}	48	-162			
D2.3	1YEH	-	CTR-----(WGFIPVREDYVMDY)W	R	H	14	12,13	K,E/E	*,N/N	L,*/D	*,*/{3}	-163	-	0.1		
D2.3	1YEC	+	CTR-----(WGFIPVREDYVMDY)W	R	H	14	12,13	K,E/E	*,N/N	L,*/D	*,*/{3}	-168	-			
D2.5	1YEE	+	CVR-----(WGFIPVREDYVLDY)W	R	Y	14	12,13	K,E/E	*,N/N	L,*/D	*,*/{3}	-167	-			
137-15A2	1AQK	+	CAR-----(VLFQQLVLYAPFDI)W	L	F	14	12	K/K	*/T	D/L:N2	*/{8}	43	-109			
6D9	1HYY	+	CAR-----(VSHYDGSRDWYFDV)W	L	Y	14	12	K/K	*/G	*/L:N2	*/{8}	42	-92			
F11.2.32	1MF2	-	CAR---(SGGIERYDGTYYVMDY)W	L	Y	16	14	K/K	*/T	*/L:C6	T/T:I	38	-107		0.3	
F11.2.32	2HRP	+	CAR---(SGGIERYDGTYYVMDY)W	L	Y	16	14	K/K	*/G	*/L:C6	T/T:I	38	-96	0.8	0.4	
17B	1GC1	+	CAG(VYEGEADEGEYDNNGFLKH)W	L	Y	19	17	K/K	*/T	L/D	*/{2}	44	-120			

[1]The symbols + and − show a complexed crystal structure with an antigen and a free structure, respectively.
[2]The sequences of the CDR-H3 segments from 1 to *n* by one-letter amino acid codes surrounded by parentheses, with the three preceding and one following residues. L46 and L49 show the amino acid residues at positions 46 and 49 of the light chain, respectively. Red and blue normal letters are the key residues for rules from i-a to i-d to identify kinked and extended bases, respectively. Red and blue italic letters are for rules from ii-a to ii-b, and rule ii-c, respectively. Olive and green letters are for rules from iii-a to iii-b, and rule iv, respectively. The notable signals are emphasized by red outlines.
[3]The length *n* of the CDR-H3 segments [13].
[4]The length *m* of the predicted β-hairpin in the CDR-H3 segments.
[5]Comparison of the predicted and crystal structures. Incorrect predicted structures are underlined. When the notable signal is observed, an alternative structure is also shown. When no meaningful feature is observed, the symbol * is indicated.
[6]K and E shows the kinked and extended bases, respectively.
[7]The symbols +, T, G, and C correspond to the K^{+}, K^{T}, K^{G}, and K^{C} stem conformations with the kinked bases, respectively. P and N are E^{P} and E^{N} structures with the extended bases, respectively.
[8]L shows a β-hairpin with a typical hydrogen bond ladder. The farthest hydrogen bond from the base as N*i* or C*i* (*i* = 2, 4, or 6) defined in Fig. 2a are also shown in parentheses. D shows a deformed β-hairpin structure without a typical hydrogen bonding ladder.
[9]T shows a typical β-turn structure. In parentheses, β-turn types are indicated. I, II, III, I′, II′, and III′ correspond to type I, type II, type III, type I′, type II′, and type III′ β-turn, respectively. R and L are the right- and left-handed local helical conformations of the residue. {j} indicates that the j residues form a β-hairpin loop, which is different from any typical β-turns.
[10]A pseudo-dihedral angle formed by the successive four C^{α} atoms at the (*n*−2)nd, (*n*−1)st, (*n*)th, and (*n*+1)st residues.
[11]A pseudo-dihedral angle formed by the successive four C^{α} atoms at the (*n*−4)th, (*n*−3)rd, (*n*−2)nd, and (*n*−1)st residues. For a kinked base with less than six residues and an extended base, this value is not calculated.
[12]A: The rmsd values of the CDR-H3 backbone atoms between the antigen-free form and the complexed form. B: The rmsd values of the CDR-H3 backbone atoms, where more than one antibody molecules were determined in different chains in the crystals.
[13]Data set 1 includes the 55 antibody structures registered in PDB release #76. Data set 2 includes the additional 45 antibodies before October, 1998.
[14]An antibody with the PDB code 1VFB was used as the representative structure of the antigen bound form of D1.3 in our previous study [8]. We here replaced it with an antibody 1A2Y with the higher resolution.
[15]An antibody with the PDB code 1JEL was used as the representative structure of JEL142 in our previous study [8]. We replaced it with an antibody 2JEL with the higher resolution.

can identify a hydrogen bond ladder by rules iii-a and iii-b, and a β-turn by rule iv.

3.2. Blind test of the original rules

The original rules in the upper half of Table 1 were examined with the 45 new CDR-H3 segments registered with the PDB [11] before October, 1998. The results are summarized in Table 2, in addition to those of the 55 CDR-H3 segments studied for the PDB release #76.

Out of the 45 new structures, 37 bases were classified correctly by the original rules from i-a to i-d. The correctly identified segments do form the hydrogen bonds or salt bridges between the key residues for judgment.

In the kinked base structures, rule ii could identify the additional bulges occasionally inserted just above the base, which were called the second bulge by Morea et al. [9]. Among the 45 newly registered CDR-H3 segments, those of BR96 with and without antigen satisfy the condition of rule ii, and the second bulges are inserted in both of the CDR-H3 segments.

Rule iii-a describes that a Pro residue at position 2, 4, *m*−3, or *m*−1 should deform the β-hairpin structure, and break the hydrogen bond ladder. Among the 45 new structures, four antibodies, C219, TR1.9, B1, and 137-15A2, correspond to rule iii-a. Pro is located at position 2 in TR1.9 and B1, and is located at position (*m*−1) in C219, B1, and 137-15A2. All of the CDR-H3s have deformed β-hairpin structures, without any significant hydrogen bond ladder. In 137-15A2, the hydrogen bond between the amide of the second residue and the (*m*−1)st carbonyl, denoted as N2 (Fig. 2a), is formed. The reason why the Pro residue deforms an extended β-structure is simply due to the lack of an amide group. Therefore, Pro at position (*m*−1) does not necessarily break the N2 hydrogen bond, because the Pro residue can provide a carbonyl oxygen for this hydrogen bond. However, in 137-15A2, no distinctive hydrogen bond ladder is found in the CDR-H3 segment above the N2 hydrogen bond. Rule iii-b describes that aromatic residues located at adjacent positions on opposite strands of the kinked base tend to form hydrogen bond ladders. RFT5, 7E2, FV4155, TP7, and 17B are newly registered antibodies for which rule iii-b applies. In fact, four out of those β-hairpins, except the longest, 17B, have typical hydrogen bond ladders.

Rule iv describes that Gly, Asp, and Asn at specific positions within the β-hairpin scheme tend to form a typical β-turn. Nine antibodies, RFT5, MAB1-IA, H57, BR96 with and without antigen, A6, 184.1, and F11.2.32 with and without antigen, are newly registered and can be examined with rule iv. All of them form typical β-turns, except 184.1 with the long segment length 12. The effectiveness depending on the segment length is discussed later.

Consequently, the original rules about the CDR-H3 conformation [8] predicted most of the antibody crystal structures that were registered after our initial proposal. When the original rules were found, their physical origins were carefully inspected. In the below sections, we again examined the exceptional cases and proposed additional rules.

3.3. Notable signals for base identification

When we first determined the original rules, one CDR-H3 fragment in YST9.1 was an exception of rule i-a, and those of 26-10 in the antigen-free and complexed forms were also exceptions of rule i-b [8]. Similarly, eight fragments out of the newly registered 45 structures were erroneously classified by rules i-b and i-c. However, interesting features are found for those cases.

The five bases in M41, ME36.1, 39-A11 with and without antigen, and E5.2, are kinked, in contrast to rule i-b. The side chains of the (*n*−1)st Asp of ME36.1 and 39-A11 form salt bridges with the first Lys and the fifth Arg in the CDR-H3 loops, respectively. The (*n*−1)st Asp in M41 forms a salt bridge with the L49-Lys. The characteristic hydrogen bonds are then formed between the backbone carbonyl oxygens of

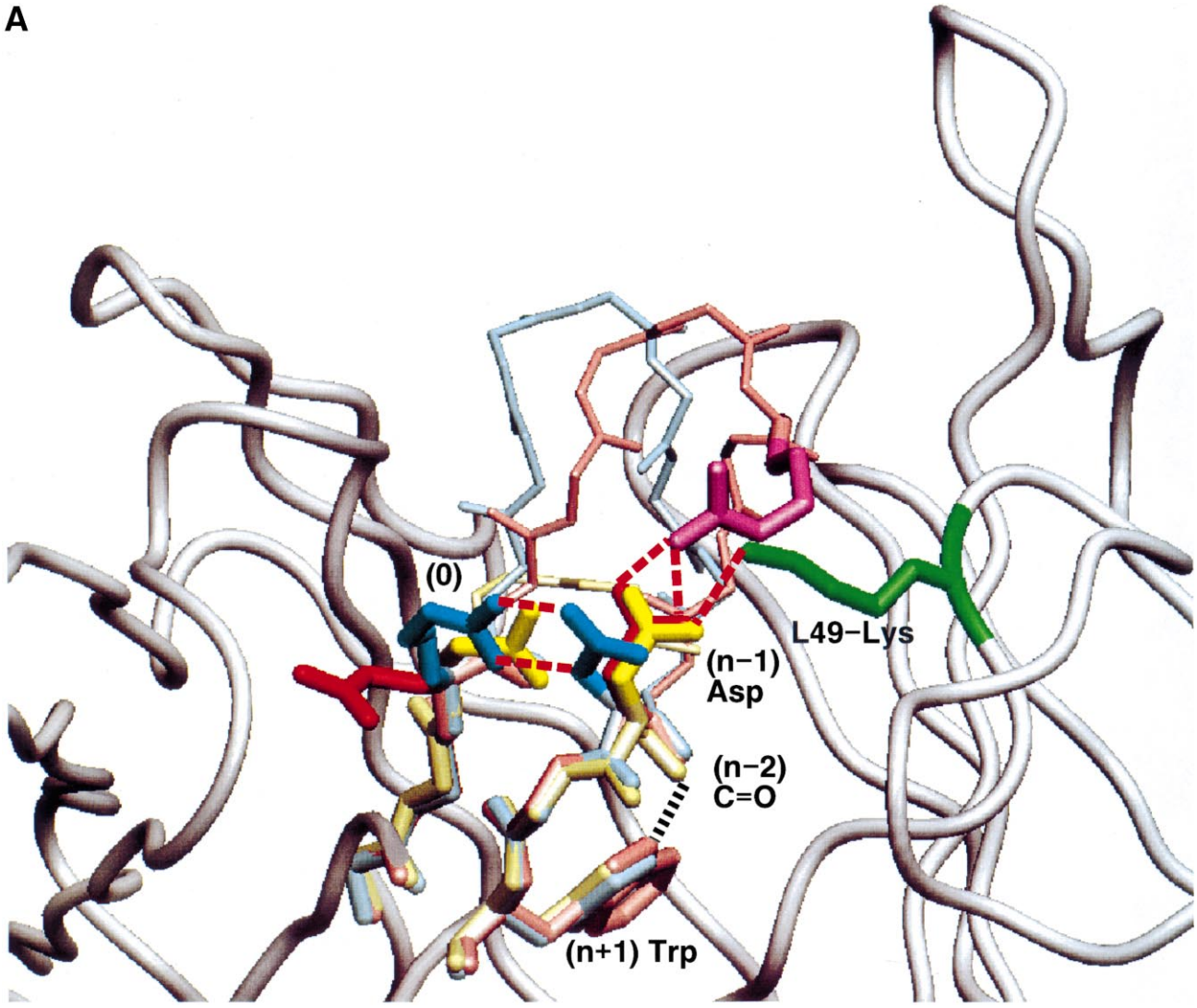

Fig. 1. Two different base structures in the CDR-H3 segments. The thick and thin lines are the base and β-hairpin structures, respectively. The red and the black dotted lines are salt bridges and hydrogen bonds, respectively. A: The kinked base in DESIRE-1 (PDB code, 1KB5) is light blue, that in M41 (1GPO) is light yellow, and that in 39-A11 (1A4J) is light red. The white pipe model shows the backbone of M41 surrounding the CDR-H3. The residues at positions 0 and ($n-1$) are colored blue in DESIRE-1, yellow in M41, and red in 39-A11, respectively. The fifth Arg in 39-A11 and the Lys at L49 in M41 are colored pink and green, respectively, as the notable signals. B: The extended base of the E^P class in MN12H2 (1MPA) is light blue, that in the E^N class in 8F5 (1A3R) is light yellow, and that in the E^N class in D2.3 (1YEC) is light red. The white pipe model shows the backbone of D2.3 surrounding the CDR-H3. The residues at positions 0 and ($n-1$) are colored blue in MN12H2, yellow in 8F5, and red in D2.3, respectively. The C^α atoms of the (0)th and the ($n-2$)nd residues are indicated by spheres. The Arg at L46 in D2.3 is colored green as the notable signal.

the ($n-2$)nd residues and the side chains of the ($n+1$)st Trp, and the typical kinked base conformations appear, as in the previous exceptional case of 26-10. Thus, when the sequence satisfies the condition of rule i-b, we must pay attention to the basic residue between the first and the ($n-3$)rd position in the CDR-H3 and to position L49 (Fig. 1A). In E5.2, the ($n-1$)st Asp is flipped to form a hydrogen bond with the fifth Tyr side chain instead of the first Lys or the eighth Arg, and the kinked base structure is formed.

The three bases in D2.5 and D2.3 in the antigen-free and complexed forms are extended, in contrast to rule i-c. In each structure, the side chain of the ($n-1$)st Asp makes a salt bridge with the Arg residue at position L46, instead of the salt bridge with the (0)th basic residue that is typical for the kinked base. In most antibodies, the position at L46 is occupied by Leu (Table 2). So, when the sequence is classified by rule i-c, the basic residue at position L46 should be a signal to form an extended base against the rule (Fig. 1B).

Among the 11 erroneously predicted bases in the total of 100 antibodies in Table 2, Lys or Arg residues at specific positions flip the ($n-1$)st Asp to form salt bridges in nine of the antibodies, in contrast to the previously proposed rules i-b and i-c. We call these basic residues at specific positions the notable signals, which are denoted by red emphasized letters in Table 2.

It should be noted that 86 out of the total of 87 CDR-H3 structures that lack the notable signals can be correctly classified only from the sequences. The one exception is a short fragment in YST9.1, which is composed of eight residues with two Pro residues. The effect of the Pro is probably stronger

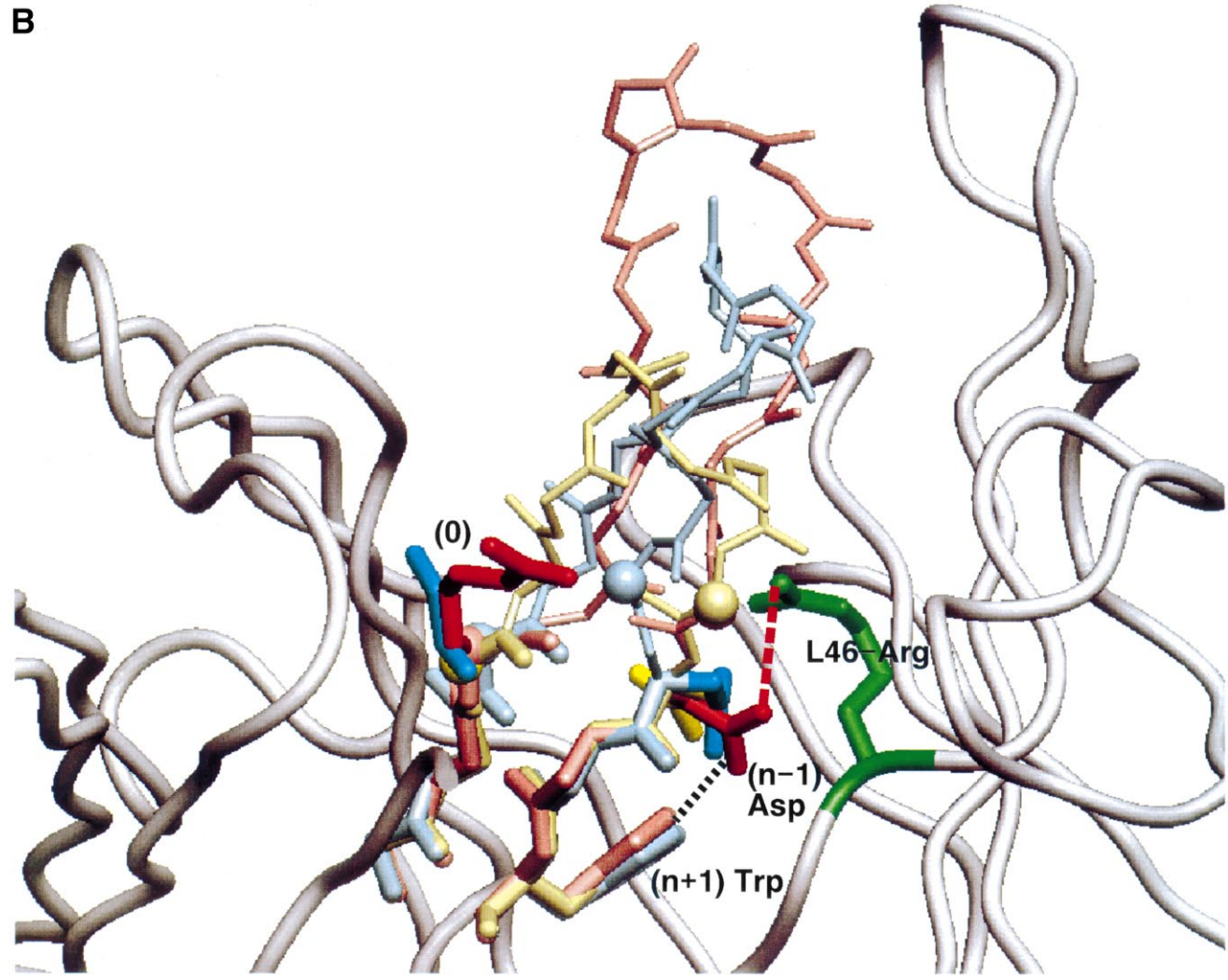

Fig. 1 (*continued*).

than the effect of the base formation for relatively short CDR-H3 segments, and the two Pro residues in YST9.1 could greatly deform the β-hairpin structure, as it obeys rule iii-a.

Alternatively, nine fragments out of the 13 CDR-H3s with the notable signals have the above features. However, four fragments, in JEL103 with and without antigen, RFT5, and E5.2, are not governed by the notable signals.

Consequently, 95 (86+9) out of the 100 bases can be correctly classified when both the original and the additional rules are applied.

3.4. Classification of the kinked bases

In our recent study of enhanced conformational sampling of the CDR-H3 fragments [13] using a multicanonical molecular dynamics simulation [14], we found that the Gly residue at position ($n-3$) in the CDR-H3 has a critical role in forming a characteristic backbone conformation just above the kinked base. Namely, the pseudo-dihedral angle formed by the four successive C^{α} atoms at the ($n-1$)st, ($n-2$)nd, ($n-3$)rd, and ($n-4$)th residues, denoted Θ_{stem}, is around $-45°$, with the (φ, ψ) backbone angle of the Gly residue at the irregular region termed ε by Efimov [15].

This tendency was also examined for the total of 81 fragments with the kinked bases and with more than six residues. There are three different stem conformations, depending upon the Θ_{stem} values: K^T (*trans*, $-170° < \Theta_{stem}, \leq -100°$), K^G (*gauche*, $-100° < \Theta_{stem}, \leq -10°$), and K^C (*cis*, $-10° < \Theta_{stem}, \leq 50°$). The 10 CDR-H3 fragments with Gly at position ($n-3$) always assume the K^G forms. Our rule ii, about the insertion of the additional bulge, also describes the same local backbone conformation. Therefore, we formalized rule ii-a in Table 1.

In addition, when the second bulge is inserted in the six K^+ forms, SE155-4, BV04-01 with and without antigen, BR96 with and without antigen, and R45-45-11, Θ_{stem} assumes only the *cis* conformations without exceptions. Thus, this feature can be another rule ii-b, as described in Table 1.

There are 65 other CDR-H3 segments with the kinked bases and with more than six residues, which are identified by neither rule ii-a nor ii-b. Among the 65 segments, most of the Θ_{stem} values are those of K^T forms in the 48 fragments.

3.5. Classification of the extended bases

Morea et al. pointed out that the extended base structures could be classified into two forms, one class like 4-4-20 and the other like HyHEL5 [9]. As indicated in Fig. 1 of our

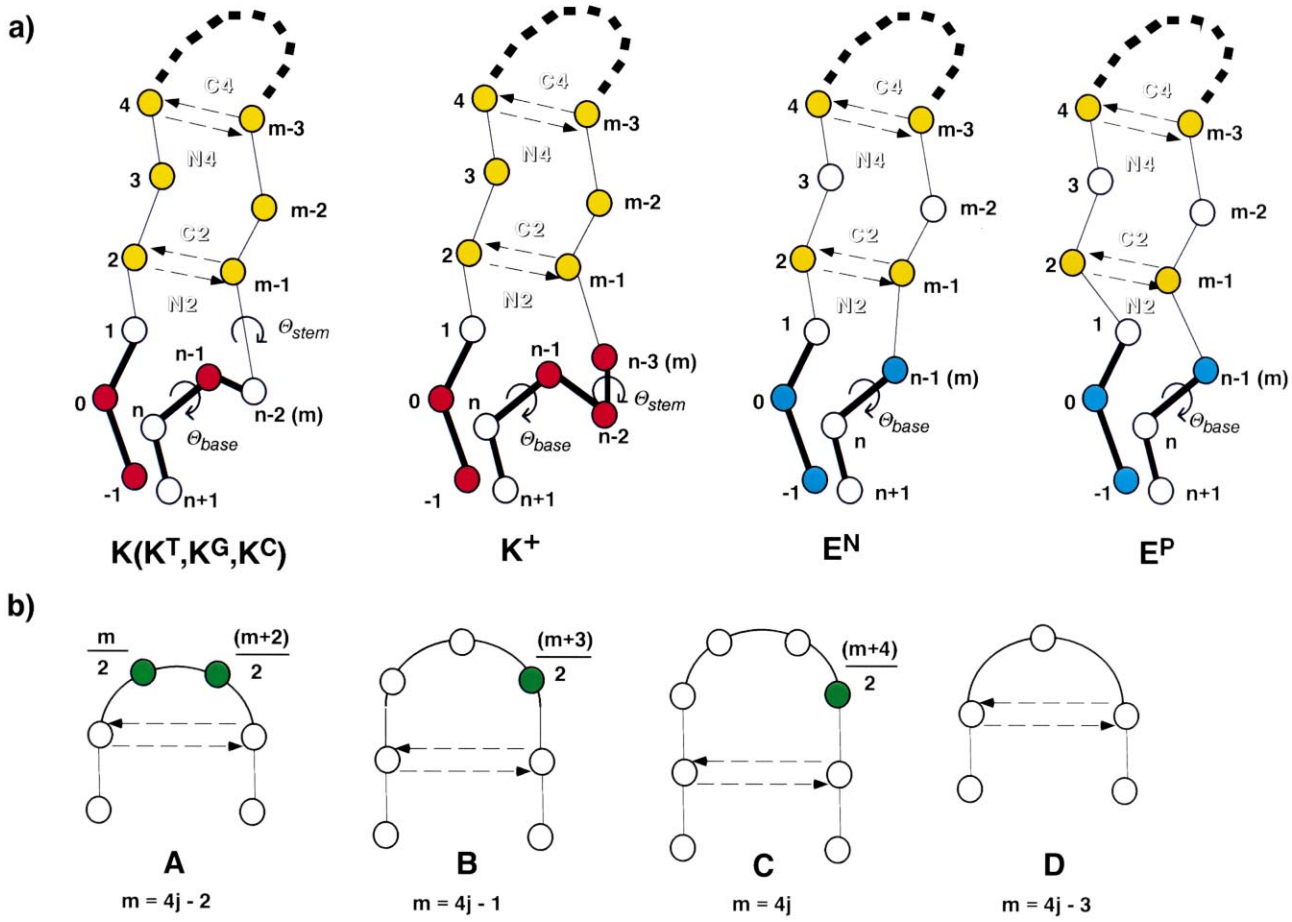

Fig. 2. Schematic representation of the CDR-H3 segment. a: Shown from left to right: a standard kinked base (K) without the additional second bulge, a kinked base with the additional second bulge (K^+), an extended base E^N with a negative Θ_{base} value, and an extended base E^P with a positive Θ_{base} value. The backbone is drawn as linked circles representing the C^α atoms with the residue numbers. The letters n and m are the numbers of the CDR-H3 segment and the β-hairpin, respectively. The hydrogen bonds between the strands are indicated by the broken arrows that point directly from the amide to the carbonyl group. N2 and C2 show the hydrogen bonds between at residues at positions 2 and $m-1$. N4 and C4 show the hydrogen bonds between at residues of positions 4 and $m-3$. Red and blue circles are residues to identify the kinked and extended bases, respectively. Yellow circles are for identifying the formation or deformation of the hydrogen bond ladder. b: Four classes of β-hairpins from A to D, following Sibanda et al. [12]. Green circles represent the key residue positions to identify the tendency to form β-turns by rule iv. j is an integer larger than zero.

previous article [8] and in Table 2, the pseudo-dihedral angle formed by the four successive C^α atoms at the $(n-2)$nd, $(n-1)$st, (n)th, and $(n+1)$st residues, denoted as Θ_{base}, is obviously a good measure to discriminate the two forms of the extended bases (see Fig. 1B). We denote the former class as the E^N with the Θ_{base} value around $-160°$, and the latter class as the E^P with the Θ_{base} value around 160°.

Only Gly or Arg residues appear at position (0) of the E^N forms. In the E^P form, moderately sized side chains, but neither Gly nor Arg, appear at position (0). These side chains interdigitate with the side chains of the $(n-2)$nd residue and pull the C-terminal strand of the CDR-H3. In contrast, when Gly is located at position (0), it is too small to interact with the $(n-2)$nd side chain in the E^P form. When the (0)th residue is Arg, the side chain would make a steric clash with the side chain of the residue at position $(n-2)$ in the E^P form. Thus, we may identify these two classes from the sequence with the new rule ii-c, as indicated in Table 1. The only exception of this rule ii-c is YST9.1 with a very deformed CDR-H3 fragment including two Pro residues.

3.6. H3-rules

The overall relationships between the sequence and the conformation in the CDR-H3 segments, including the notable signals and the new rules ii-a, -b, and -c, are summarized into a rule set denoted as the 'H3-rules', as indicated in Table 1. The original and the additional rules are in the upper and the lower half, respectively. They can provide identification of the CDR-H3 conformations from only the amino acid sequences, and they should be useful when antibody structural models are constructed [16,17].

3.7. Dependence on the fragment length

The CDR-H3 fragments with a length of five residues must be the shortest to construct the kinked bases. A total of seven fragments with lengths of five residues with the kinked bases were analyzed, and all of them, except M41, have the type II β-turn plus the α_L conformation [12] on the kinked base.

Some relationships are notable between the length and the availability for rule iv. In the total of 49 structures with CDR-H3 fragments longer than six residues and shorter than or

equal to ten residues, 16 CDR-H3s (32.6%) are applicable for rule iv, and 15 of the 16 fragments (93.8%) actually form the typical β-turns. On the contrary, in the total of 44 structures with more than 10 residues, six CDR-H3s (13.6%) correspond to rule iv, and only three out of six fragments (50.0%) form the β-turns.

Consequently, when the length of the CDR-H3 loop is from six to 10 residues, the amino acid sequences of many loops satisfy the condition of rule iv, and the sequence-structure relationship described by rule iv agrees with nearly 90% accuracy. In contrast, for the CDR-H3 fragments longer than 10 residues, rule iv does not work at all.

3.8. Structural flexibility of CDR-H3

Some of the CDR-H3 segments are so flexible that significant conformational changes appear as a consequence of antigen binding [5]. In contrast, many antibodies are rather rigid, and only small changes are observed upon antigen binding. In order to capture the molecular recognition scheme more precisely, it is necessary to recognize both the antigen-free and complexed forms of an antibody.

Table 2 indicates the backbone rmsds of the CDR-H3 conformational changes upon antigen binding and those of the CDR-H3s among the different chains in the crystals. From these rmsd values, it is evident that significant structural changes with rmsds larger than 1 Å occur only in the CDR-H3 segments with β-hairpin amino acid sequences that lack the features for both rules iii-b and iv. In contrast, when the sequences satisfy the condition of either rule iii-b or iv, structural changes with rmsds smaller than 0.8 Å are observed upon antigen binding, except for the complexation of F11.2.32. Even in this exceptional case, the change is not very large, considering its long CDR-H3 fragment. Therefore, we propose an empirical relationship when the sequence satisfies the condition of either rule iii-b or iv, the β-hairpin structure is not flexible and does not show any large structural change upon antigen binding. This tendency encourages us to make a complex model when the CDR-H3 fragments are considered to be rigid.

3.9. Conclusion

To identify the tertiary structure from the amino acid sequence of the CDR-H3 segment of an antibody, the H3-rules were described. For enhancement of the affinities and the selectivities of antibodies, the CDR-H3 segment is an attractive target for its functional potential by the combinatorial approach. In fact, an attempt to improve the activity of an antibody by in vitro affinity maturation has been performed [18]. It is certainly an effective approach, but the number of changeable residues is still limited to several residues. Therefore, the current H3-rules can accelerate the screening procedure to identify the target residue positions in the CDR-H3 [19].

References

[1] Wilson, I.A. and Stanfield, R.L. (1994) Curr. Opin. Struct. Biol. 4, 857–867.
[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. Dept. HHS, NIH Publication No. 91-3242, Bethesda, MD.
[4] Holliger, P. and Hoogenboon, H. (1998) Nature Biotechnol. 16, 1015–1016.
[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] Al-Lazikani, B., Lesk, A.M. and Chothia, C. (1997) J. Mol. Biol. 273, 927–948.
[7] Stanfield, R.L. and Wilson, I.A. (1994) Trends Biotechnol. 12, 275–279.
[8] Shirai, H., Kidera, A. and Nakamura, H. (1996) FEBS Lett. 399, 1–8.
[9] Morea, V., Tramontano, A., Rustici, M., Chothia, C. and Lesk, A.M. (1998) J. Mol. Biol. 275, 269–294.
[10] Oliva, B., Bates, P.A., Querol, E., Aviles, F.X. and Sternberg, M.J.E. (1998) J. Mol. Biol. 279, 1193–1210.
[11] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[12] Sibanda, B.L., Blundell, T.L. and Thornton, J.M. (1989) J. Mol. Biol. 206, 759–777.
[13] Shirai, H., Nakajima, N., Higo, J., Kidera, A. and Nakamura, H. (1998) J. Mol. Biol. 278, 481–496.
[14] Nakajima, N., Nakamura, H. and Kidera, A. (1997) J. Phys. Chem. B101, 817–824.
[15] Efimov, A.V. (1986) Mol. Biol. (Moscow) 20, 250–260.
[16] Rees, A.R., Searle, S.J., Henry, A.H., Whitelegg, N. and Pedersen, J. (1996) Protein Structure Prediction: A Practical Approach (Sternberg, M.J.E., Ed.), pp. 141–172, Oxford University Press, Oxford.
[17] Mandal, C., Kingery, B.D., Anchin, J.M., Subramaniam, S. and Linthicum, D.S. (1996) Nature Biotechnol. 14, 323–328.
[18] Fujii, I., Fukuyama, S., Iwabuchi, Y. and Tanimura, R. (1998) Nature Biotechnol. 16, 463–467.
[19] Any CDR-H3 amino acid sequence can be inspected in our home page: http://www.beri.co.jp/~nakamura/H3-rules.html.