**Chothia Rules**

In terms of annotation (without considering methods and concepts) the Chothia annotation is similar to the Kabat annotation, excepting small changes in some CDRs boundaries and changes in the insertion positions. So the better way to get the Chothia annotation – just to use the Kabat Rules (they easily work on sequences) and then add some necessary changes (whose can be found here below) to transform it into the required Chothia annotation. (Because the Chothia original main rules does not exist in the way the Kabat ones does. Kabat presented his main rules separately and clearly unlike Chothia’s team – the Chothia main rules can be obtained only by pieces along all their articles, so they are a bit artificial.)

The best way to get the most accurate linguistic annotation is to use **ICK-Data** (based on Kabat R, Chothia R and IMGT R) so check and study ***“ICK-Data about.doc”*** and ***“ICK-Data.xls”***.

**Light chain**

FR1 – **ends *nearly* after Cys23**, length = 25 for V(k) and 24 for V(lambda), (1-25) for V(k) and (1-24) for V(lambda)

CDR1 = L1 – **begins *nearly* after Cys23 and ends *nearly* before Trp35**, length <= 13, (26-32) for V(k) and (25-32) for V(lambda) ///// 30a, b, c, d, e, f

Note that Chothia’s team were using incorrect 31-insertion-position in some of their articles => but in the latest articles it was changed to the correct 30-insertion-position //so if some of considered articles use the incorrect 31-insertion-position try *to translate it into the correct one* or *don’t use that info*.

FR2 – **begins *nearly* before Trp35** (and ends *nearly* after Ile48), length = 17, (33-49)

CDR2 = L2 – (begins *nearly* after Ile48), length <= 3, (50-52)

FR3 – **ends *nearly* after Cys88**, length = 38, (53-90)

CDR3 = L3 – **begins *nearly* after Cys88 and is followed by Val/Thr/Gly97**, length <= 8 (12), (91-96) ///// 95a, b, c, d, e, f

FR4 – **begins with Val/Thr/Gly97**, (nearly invariant Gly99 and Gly101), length = 11-12, (97-107) ///// 106A

**Heavy chain**

FR1 – **ends *nearly* after Cys22 = is followed by Gly26**, (Ser/Val/Trp/Ala24), length = 25, (1-25)

CDR1 = H1 – **begins *nearly* after Cys22 = starts with Gly26 and ends *nearly* before Trp36**, length <= 9, (26-32) ///// 31a, b

FR2 – **begins *nearly* after Trp36** (and ends with Ile/Val/Ser51), length = 20, (32-51)

CDR2 = H2 – (begins after Ile/Val/Ser51), length <= 8, (52-56) ///// 52a, b, c

FR3 – **ends *nearly* after Cys92**, length = 29-32, (57-94) ///// 82a, b, c

CDR3 = H3 – **begins *nearly* after Cys92 and ends *nearly* before Gly104**, length <= 19, (95-102) or (96-101) ///// 100a, b, c, d, e, f, g, h, i, j, k

FR4 – **begins nearly before Gly104**, (Gly106), length = 11, (103-113)

***Numbering:***

**Light-chain:** 1 … 30 30a 30b 30c 30d 30e 30f 31… 94 95 95a 95b 95c 95d 95e 95f 96 … 105 106 106a 107…

**Heavy-chain:** 1 … 31 31a 31b 32… 51 52 52a 52b 52c 53 … 81 82 82a 82b 82c 83 … 99 100 100a 100b 100c 100d 100e 100f 100g 100h 100i 100j 100k 101…

(… - numbering equals to the sequence of natural numbers)

**Afterword:** CDR = hypervariable region / FR – **information about (mostly) invariable key residues that helps to find this region (if they exist),** **(**information about *less* invariable residues but also possible for use (if they exist)**),** length of the region (that help to move from the key residues and find the regions)**,** **(x-y)** – region’s boundaries corresponding to the Kabat numbering (can help if the numbering already exists or to check the already annotated sequence) **/////** addition positions in the numbering, that are located in this region (can help with making the correct numbering)

// information marked with a gray color supposed to be correct, but don’t have a proof in the considered articles

*<< 6, 1987 >> residues buried within VL and VH domains (mostly invariant ones)*

*<< 4, 12, 1987 >> boundaries (96-101?) and add-positions for H3*

*<< 5, 1992 >> conserved Gly26 in VH*

*<< 11, 1992 >> add-positions and numbering for H1, H2*

*<< 3, 1995 >> full VL sequence numbering (but incorrect 31-insertion-position for L1)*

*<< 4, 1995 >> add-positions and numbering for L1 V(k) (but incorrect 31-insertion-position for L1)*

*<< 4, 6, 7, 1995 >> VL CDRs boundaries*

***<< 1997 >> V(k) and V(lambda) L1, L2, L3 and VH H1, H2 boundaries, gaps and numberings (correct 30-insertion-position for L1!):***

***3 - L1 V(k) boundaries and numbering***

***8 - L1 V(lambda) boundaries and numbering***

***9-10 - L2 boundaries and numbering***

***9-11 - L3 V(K) boundaries and numbering***

***9, 12 - L3 V(lambda) boundaries and numbering***

***17, 13 - H1 boundaries and numbering***

***17, 14-15 - H2 boundaries and numbering***

***18 - about H3 boundaries (95-102?)***

**Extra (mostly and nearly invariant residues)**

**See below about different canonical structures and their key residues (in the section after next)**

**// not here!**

**Residues commonly buried within VL and VH domains: *<< 6, 1987 >>***

**// i.e. mostly and nearly invarinat residues**

**VL:** 4 LM, 6 Q, 19 V, 21 IM, 23 C, 25 GAS, 33 VL, 35 W, 37 Q, 47 LIW, 48 I, 62 F, 64 GA, 71 AFY, 73 LF, 75 IV, 82 D, 84 AS, 86 Y, 88 C, 90 ASQN, 97 VTG, 99 G, 101 G, 102 T, 104 LV

**VH: 4 L, 6 QE, 18 L, 20 L, 22 C, 24 SVTA, 34 MY, 36 W, 38 R, 48 IV, 49 AG, 51IVS, 69 IVM, 78 LF, 80 L, 82 ML, 86 D, 88 AG, 90 Y, 92 C, 104 G, 106 G, 107 TS, 109 V**

**26 G – absolutely conserved in VH *<< 5, 1992 >>***

**More accurate for some CDRs:**

**L1 V(lambda)**: usually 71 A, 90 SA *<< 3, 1987 >>*

*Human:* 25 G, 30 V(I), 33 V(I), 71 A, 29 DN(S) *<< 6, 1987 >>*

**L1 V(k)**: usually 71 YF, 90 QN *<< 3, 1987 >>*

***Human:* 2 I(MV), 25 AS, 29 IVL, 33 L(V), 71 F(Y)**

***Mouse:* 2 IV, 25 A(S), 29 LVI, 33 LMV(I), 71 FY** *<< 7, 1987 >>*

**L2**: **48 I, 64 GA *<< 8, 1987 >>***

**L3 V(k): usually 90 QN, 95 P *<< 10, 1987 >>***

**H1: 26 G, 27 FT, 29 F, 34 M(I), 94 RK *<< 10, 1987 >>***

***(Full) JH segments for Human and Mouse: << 12, 1987 >>***

***HUMAN***

JH1: AEYFGHWGQGTLVTVSS

JH2: YWYFDLWGRGTLVTVSS

JH3: AFDVWGQGTMVTVSS

JH4: YFDYWGQGTLVTVSS

JH5: NWFDSWGQGTLVTVSS

JH6: YGMDVWGQGTTVTVSS

**MOUSE**

JH1: WYFDVWGAGTTVTVSS

JH2: YFDVWGQGTTVTVSS

JH3: WFAYWGWGTLVTVSA

JH4: DYWGWGTSVTVSS

**Chothia CDRs Boundaries along the years**

**1987:** *<< 4, 1987 >>*

26-32, 50-52, 91-96 – L1, L2, L3

26-32, 53-55, 96-101 – H1, H2, H3

**1989:**  
26-32, 50-52, 91-96 – V(k) L1, L3, L3 ***<< 2, 1989 >>***  
26-32, 52a-55 – H1, H2 ***<< 3, 1989 >>***

**1992:**   
26-32, 52-56 – H1, H2 ***<< 3, 5, 1992 >>***

**1995:**  
26-32, 50-52, 91-96 – V(k) L1, L2, L3 ***<< 4, 1995 >>***

**1997:**  
26-32, 50-52, 91-96 – V(k) L1, L2, L3 ***<< 3, 9, 1997 >>***25-32, 50-52, 91-96 – V(lambda) L1, L2, L3 ***<< 8, 9, 1997 >>***26-32, 52-56 – H1, H2 ***<< 17, 1997 >>***  
95-102 ??? – H3 ***<< 18, 1997 >>***

**Chothia canonical structures along the years**

**Chothia 1989**

**Canonical structures**

**V(k) *<< 2, 1989 >>***

**L1 (26-32) (95 human, 299 mouse)**

1. **Canonical structure L1-1 // 0% human, 15% mouse**

**26 S, 27 S, 28 S, \*29 V, 30 SNR, 31 -, 31a -, 31b -, 31c -, 31d -, 31e -, 31f -, 32 SY // \*2 I, \*25 A, \*33 LM, \*71 Y**

**///// SSSVSS (I-A-L-Y), SSSVNY (I-A-M-Y), SSSVRY (I-A-M-Y)**

1. **Canonical structure L1-2 // 60% human, 25% mouse**

**26 S, 27 QG, 28 DNS, \*29 IV, 30 IHGS, 31 KNT, 31a -, 31b -, 31c -, 31d -, 31e -, 31f -, 32 VYNA // 2 I, 25 A, 33 L, 71 YF**

**///// SQDIIKV (I-A-L-Y), SGNIHNY (I-A-L-Y), SQSIGNN (I-A-L-F), SQDVSTA (I-A-L-Y)**

1. **Canonical structure L1-3 // 5% human, 20% mouse**

**26 S, 27 E, 28 S, \*29 L, 30 L, 31 N, 31a S, 31b G, 31c N, 31d E, 31e K, 31f N, 32 F // 2 I, 25 S, 33 L, 71 F**

**///// SESLLNSGNEKNF (I-S-L-F)**

1. **Canonical structure L1-4 // 5% human, 10% mouse**

**26 S, 27 Q, 28 S, \*29 L, 30 V, 31 H, 31a S, 31b -, 31c N, 31d G, 31e N, 31f T, 32 Y // 2 V, 25 S, 33 L, 71 F**

**///// SQSLVHSNGNTY (V-S-L-F)**

**L2 (50-52) (69 human, 183 mouse)**

1. **Canonical structure L2-1 // 95% human, 95% mouse**

**50 EGYDWK, 51 AITV, 52 ST // \*48 I, \*64 G**

**///// EAS, GAS, EIS, YTT, DTS, YAS, WAS, DTS, KVS (I-G for all)**

**L3 (91-96) (52 human, 152 mouse)**

1. **Canonical structure L3-1 // 90% human, 80% mouse**

**91 YDFSHW, 92 QHWNYTS, 93 SH, 94 LYTWPVN, \*95 P, 96 YLRW // \*90 QNH**

**///// YQSLPY (Q), DHSYPL (N), FWSTPR (H), SNSWPY (Q), HYSPPW (Q), STHVPW (Q), WSSNPL (Q)**

1. **Canonical structure L3-2 // 0% human, 10% mouse**

**91 W, 92 T, 93 Y, \*94 P, 95 L, 96 I // \*90 Q**

**///// WTYPLI (Q)**

1. **Canonical structure L3-3 // 2% human, 1% mouse**

**91 W, 92 G, 93 R, 94 N, \*95 P, 96 - // \*90 Q**

**///// WGRNP (Q)**

**VH *<< 3, 1989 >>***

**H1 (26-32) (50 human, 321 mouse)**

1. **Canonical structure H1-1 // 50% human, 80% mouse**

**\*26 G, \*27 FY, 28 TIDS, \*29 FL, 30 ST, 31 DSKGN, 32 FY // \*34 MVI, \*94 RG**

**///// GFTFSDF (M-R), GFIFSSY (M-R), GFDFSKY (M-R), GFSLTGY (V-R), GYTFSDY (I-R), GYTFTNY (M-R), GFTFSSF (M-R), GFTFSDY (M-G)**

1. **Canonical structure H1-1’ // also the same as H1-1**

**26 G, 27 SD, 28 TS, 29 FI, 30 ST, 31 ND, 32 D // 34 YW, 94 RN**

**///// GSTFSND (Y-R), GDSITDD (W-N)**

**H2 (52a-55) (54 human, 248 mouse)**

1. **Canonical structure H2-1 // 15% human, 15% mouse**

**52a -, 52b -, 52c -, 53 YG, 54 HDS, \* 55 G**

**///// YHG, GDG, YSG**

1. **Canonical structure H2-2 // 1% human, 40% mouse**

**52a PT, 52b -, 52c -, 53 GN, 54 ST, \* 55 G // \* 71 AL**

**///// PGSG (A), TNTG (L)**

1. **Canonical structure H2-3 // 40% human, 5% mouse**

**52a DPS, 52b -, 52c -, 53 DG, 54 \*SG, 55 SG // \* 71 R**

**///// DDGS (R), PDSG (R), SGSS (R)**

1. **Canonical structure H2-4 // 15% human, 20% mouse**

**52a N, 52b K, 52c GP, 53 NY, 54 \*KN, 55 \*Y // \*71 R**

**///// NKGNKY (R), NKPYNY (R)**

**\* - “The residues in the hypervariable and framework regions that are mainly responsible for these conformations are indicated by asterisk” *<< 3, 1989 >>***

**Chothia 1992**

**Human canonical structure classes and families:**

**// “families” also can be called “subgroups”**

**// “**Sequences that have the same canonical structure for both H1 and H2 can be grouped together into “canonical structure classes”. These classes are numbered in the form **N-M where N is the number of the H1 canonical structure and M the number of the H2 structure**.” ***<< 10, 1992 >>***

**1-1 *<< 11, 1992 >>***

1. **Family 3**

**H1-1-F3 \*26 G, \*27 F, 28 T, \*29 VF, 30 S, 31 S, 31a -, 31b -, 32 NY, 33 YAD, \*34 M, 35 SH // \*24 AG, \*94 R**

**///// GFTVSSNYMS (A-R), GFTFSSYAMH (G-R), GFTFSSYDMH (A-R)**

**H2-1-F3 50 VA, 51 I, 52 YG, 52a -, 52b -, 52c -, 53 ST, 54 GA, \*55 G, 56 SGD, 57 T, 58 Y // \*71 R**

**///// VIYSGGSTY, AIGTGGGTY, AIGTAGDTY (R for all)**

1. **Family 4**

**H1-1-F4 \*26 G, \*27 G, 28 S, \*29 FVI, 30 S, 31 GS, 31a -, 31b -, 32 Y, 33 Y, \*34 W, 35 S // \*24 V, \*94 R**

**///// GGSFSGYYWS, GGSVSGYYWS, GGSISSYYWS, GGSVSSYYWS (V-R for all)**

**H2-1-F4 50 EYR, 51 I, 52 INY, 52a -, 52b -, 52c -, 53 HYT, 54 S, \*55 G, 56 S, 57 T, 58 N // \*71 V**

**///// EIIHSGSTN, EINHSGSTN, YIYYSGSTN, RIYTSGSTN (V for all)**

**1-2 *<< 12, 1992 >>***

1. **Family 1**

**H1-1-F1 \*26 G, \*27 YG, 28 T, \*29 F, 30 TS, 31 SD, 31a -, 31b -, 32 Y, 33 YAG, \*34 MI, 35 HSN // \*24 VA, \*94 TR-**

**///// GYTFTDYYMH (V-T), GGTFSSYAIS (A-R), GYTFTSYAMN (A-R), GYTFTSYGIS (A-R/A--)**

**H2-2-F1 50 LGRW, 51 VI, 52 DINS, \*52a PTA, 52b -, 52c -, 53 EINY, 54 DFLTN, \*55 G, 56 ETIN, 57 TAP, 58 INT // \*71 ALT**

**///// LVDPEDGETI (A), GIIPIFGTAN (A), RIIPILGIAN (A), WINTNTGNPT (L), WISAYNGNTN (T)**

1. **Family 5**

**H1-1-F5 \*26 G, \*27 Y, 28 S, \*29 F, 30 T, 31 S, 31a -, 31b -, 32 Y, 33 W, \*34 I, 35 GHS // \*24 GT, \*94 R**

**///// GYSFTSYWIG (G-R), GYSFTSYWIH (T-R), GYSFTSYWIS (G-R)**

**H2-2-F5 50 ISR, 51 I, 52 YD, \*52a P, 52b -, 52c -, 53 GS, 54 DN, \*55 S, 56 DY, 57 T, 58 RN // \*71 A**

**///// IIYPGDSDTR, SIYPGNSDTR, RIDPSDSYTN (A for all)**

**1-3 *<< 13, 1992 >>***

1. **Family 1**

**H1-1-F1 \*26 G, \*27 FY, 28 TI, \*29 F, 30 TN, 31 SDGNY, 31a -, 31b -, 32 SYR, 33 AYCD, \*34 VMIL, 35 QHN // \*24 A, \*94 R**

**///// GFTFTSSAVQ, GYIFTDYYMH, GYTFTGYYMH, GYTFTNYCMH, GYTFTSYAMH, GYTFTSYDIN, GYTFTSYYMH, GYTFNSYYMH, GYTFTYRYLH (A-R for all)**

**H2-3-F1 50 WRLI, 51 ISM, 52 VNCT, 52a VPA, 52b -, 52c -, 53 GNSF, \*54 SDNG, 55 G, 56 NGS, 57 T, 58 NSKG // \*71 R**

**///// WIVVGSGNTN, RINPNSGGTN, WINPNSGGTN, LVCPSDGSTS, WINAGNGNTK, WSNAGNGNTK, WMNPNSGNTG, IINPSGGSTS, WITPFNGNTN (R for all)**

1. **Family 3**

**H1-1-F3 \*26 G, \*27 F, 28 T, \*29 F, 30 DS, 31 DNS, 31a -, 31b -, 32 YH, 33 AGTYESW, \*34 MT, 35 HSN // \*24 A, \*94 RK**

**///// GFTFDDYAMH (A-K), GFTFDDYGMS (A-R), GFTFDDYTMH (A-K), GFTFSDYYMS (A-R), GFTFSNHYTS (A-K), GFTFSSYAMH (A-R), GFTFSSYAMS (A-K), GFTFSSYEMN (A-R), GFTFSSYGMH (A-K/A-R), GFTFSSYSMN (A-R), GFTFSSYWMH (A-R), GFTFSSYWMS (A-R)**

**H2-3-F3 50 GLYAVRN, 51 I, 52 SNWK, 52a WSGY, 52b -, 52c -, 53 NDS, \*54 SG, 55 GS, 56 STYNE, 57 ITK, 58 GYN // \*71 R**

**///// GISWNSGSIG, GINWNGGSTG, LISWDGGSTY, YISSSGSTIY, YSSGNSGYTN, AISSNGGSTY, VISYDGSNKY, AISGSGGSTY, YISSSGSTIY, VIWYDGSNKY, YISSSSSTIY, RINSDGSSTT, NIKQDGSEKY (R for all)**

**1-4 *<< 14, 1992 >>***

1. **Family 3**

**H1-1-F3 \*26 G, \*27 F, 28 T, \*29 F, 30 S, 31 D, 31a -, 31b -, 32 H, 33 Y, \*34 M, 35 DS // \*24 A, \*94 R**

**///// GFTFSDHYMD, GFTFSDHYMS (A-R for all)**

**H2-4-F3 50 RL, 51 TI, 52 R, 52a N, 52b K, 52c A, 53 N, \*54 S, \*55 Y, 56 T, 57 T, 58 E // \*71 R**

**///// RTRNKANSYTTE, LIRNKANSYTTE (R for all)**

**2-1 *<< 15, 1992 >>***

1. **Family 2**

**H1-2-F2 \*26 G, \*27 F, 28 S, \*29 L, 30 S, 31 T, 31a S, 31b -, 32 E, 33 W, \*34 C, 35 G // \*24 F, \*94 H**

**///// GFSLSTSEWCG (F-H)**

**H2-1-F2 50 L, 51 I, 52 Y, 52a -, 52b -, 52c -, 53 W, 54 D, \*55 D, 56 D, 57 K, 58 R // \*71 K**

**///// LIYWDDDKR (K)**

1. **Family 4**

**H1-2-F4 \*26 G, \*27 YG, 28 S, \*29 I, 30 S, 31 S, 31a GS, 31b -, 32 YN, 33 YW, \*34 W, 35 GS // \*24 V, \*94 R**

**///// GYSISSGYYWG, GYSISSSNWWG, GGSISSSNWWS (V-R for all)**

**H2-1-F4 50 SYE, 51 I, 52 Y, 52a -, 52b -, 52c -, 53 HY, 54 S, \*55 G, 56 SN, 57 TIP, 58 YN // \*71 V**

**///// SIYHSGSTY, YIYYSGSIY, YIYYSGSTY, EIYHSGSTN, EIYHSGNPN (V for all)**

**3-1 *<< 15, 1992 >>***

1. **Family 2**

**H1-3-F2 \*26 G, \*27 F, 28 S, \*29 L, 30 S, 31 NT, 31a AS, 31b RG, 32 MV, 33 GCR, \*34 V, 35 SG // \*24 VF, \*94 RH**

**///// GFSLSNARMGVS (V-R), GFSLSTSGMCVS (F-R), GFSLSTSGMRVS (F-R), GFSLSTSGVGVG (F-H)**

**H2-1-F2 50 HLR, 51 I, 52 FDY, 52a -, 52b -, 52c -, 53 SW, 54 ND, \*55 D, 56 ED, 57 K, 58 SYFR // \*71 K**

**///// HIFSNDEKS, LIDWDDDKY, RIDWDDDKF, LIYWNDDKR (K for all)**

1. **Family 4**

**H1-3-F4 \*26 G, \*27 G, 28 S, \*29 IV, 30 S, 31 S, 31a GS, 31b GS, 32 Y, 33 SY, \*34 W, 35 SG // \*24 V, \*94 R**

**///// GGSISSGGYSWS, GGSISSGGYYWS, GGSVSSGSYYWS, GGSISSSSYYWG (V-R for all)**

**H2-1-F4 50 YS, 51 I, 52 Y, 52a -, 52b -, 52c -, 53 HY, 54 S, \*55 G, 56 S, 57 T, 58 YN // \*71 V**

**///// YIYHSGSTY, YIYYSGSTY, YIYYSGSTN, SIYYSGSTY (V for all)**

**Tomlinson 1995**

**Canonical structures of Vk (addition to the Chothia’s 1989)**

**L1 (26-32)**

***<< 3, 1995 >> human Vk germlines (divided into subgroups) with the information about the canonical structures of each region***

***<< 3, 1995 >> lengths of L1 of each canonical structure***

***<< 4, 1995 >> key residues***

***<< 5-6, 1995 >> amount of sequences for each canonical structure***

***<< 6, 1995 >> conservative mutations at the key sites***

**6 canonical structures - 4 human (2, 3, 4, 6)**

1. **Canonical structure L1-2 (55% = 401)**

**// length = 7 (…30, 31, 32…)**

**// 2 I, 25 A, 29 IV, 33 LVI, 71 YF**

**// conservative mutations (small change): … to 2 FLTV, 25 GST, 29 LMT, 33 V, 71 Y**

**Subgroup I: 26 S, 27 Q, 28 SDG, \*29 I, 30 SR, 31 SN, 31a -, 31b -, 31c -, 31d -, 31e -, 31f -, 32 YDWA // \*2 I, \*25 A, \*33 L, \*71 FY**

**///// SQSISSY, SQDISNY, SQGISNY, SQGIRND, RQGISNY, SQGISSW, SQGISSA, SQGISSY, SQGISSY (I-A-L-Y), SQSISSW (I-A-L-F for all instead marked one)**

**Subgroup III: 26 S, 27 Q, 28 SG, \*29 V, 30 S, 31 S, 31a -, 31b -, 31c -, 31d -, 31e -, 31f -, 32 NY // \*2 I, \*25 A, \*33 L, \*71 F**

**///// SQSVSSN, SQSVSSY, SQGVSSY (I-A-L-F for all)**

**Subgroup V: 26 S, 27 Q, 28 D, \*29 I, 30 D, 31 D, 31a -, 31b -, 31c -, 31d -, 31e -, 31f -, 32 D // \*2 T, \*25 A, \*33 M, \*71 F**

**///// SQDIDDD (I-A-M-F)**

**Subgroup VI: 26 S, 27 QE, 28 SG, \*29 I, 30 G, 31 SN, 31a -, 31b -, 31c -, 31d -, 31e -, 31f -, 32 SY // \*2 IV, \*25 A, \*33 L, \*71 F**

**///// SQSIGSS (I-A-L-F), SEGIGNY (V-A-L-F)**

1. **Canonical structure L1-3 (7% = 54)**

**// length = 13 (…30, 31, 31a, 31b, 31c, 31d, 31e, 31f, 32…)**

**// 2 I, 25 S, 29 LV, 31d QSE, 33 L, 71 F**

**// conservative mutations (small change): … to 2 FLTV, 25 AT, 29 IMT, 33 FV, 71 Y**

**Subgroup II: 26 S, 27 Q, 28 S, \*29 L, 30 L, 31 D, 31a S, 31b D, 31c D, 31d G, 31e N, 31f T, 32 Y // \*2 I, \*25 S, \*33 L, \*71 F**

**///// SQSLLDSDDGNTY (I-S-L-F)**

**Subgroup IV: 26 S, 27 Q, 28 S, \*29 V, 30 L, 31 Y, 31a S, 31b S, 31c N, 31d N, 31e K, 31f N, 32 Y // \*2 I, \*25 S, \*33 L, \*71 F**

**///// SQSVLYSSNNKNY (I-S-L-F)**

1. **Canonical structure L1-4 (9% = 65)**

**// length = 12 (…30, 31, 31a, 31c, 31d, 31e, 31f, 32…)**

**// 2 VL, 25 SP, 29 LI, 31d G, 33 LF, 71 F**

**// conservative mutations (small change): … to 2 I, 25 AT, 29 MV, 33 V, 71 Y**

**Subgroup II: 26 S, 27 Q, 28 S, \*29 L, 30 VL, 31 YH, 31a S, 31b -, 31c DN, 31d G, 31e NKY, 31f TN, 32 Y // \*2 VI, \*25 S, \*33 L, \*71 F**

**///// SQSLVYSDGNTY (V-S-L-F), SQSLLHSDGKTY (I-S-L-F), SQSLLHSNGYNY (I-S-L-F), SQSLVHSDGNTY (I-S-L-F)**

1. **Canonical structure L1-6 (29% = 216)**

**// length = 8 (…30, 31, 31a, 32…)**

**// 2 N, 25 A, 29 V, 33 L, 71 Y**

**// conservative mutations (small change): … to 25 GST, 29 ILMT, 33 FV**

**Subgroup III: 26 S, 27 Q, 28 S, \*29 V, 30 S, 31 S, 31a S, 31b -, 31c -, 31d -, 31e -, 31f -, 32 Y // \*2 I, \*25 S, \*33 L, \*71 F**

**///// SQSVSSSY (I-S-L-F)**

**L2 (50-52)**

***<< 3, 1995 >> human Vk germlines (divided into subgroups) with the information about the canonical structures of each region***

***<< 4, 6, 1995 >> length of the L2 canonical structure***

***<< 4, 1995 >> key residues***

***<< 6, 1995 >> amount of sequences for each canonical structure***

***<< 6, 1995 >> conservative mutations at the key sites***

**1 canonical structure**

1. **Canonical structure L2-1 (99% = 736)**

**// length = 3**

**// 48 IV, 64 G // conservative mutations (small change): … to 48 LMV, 64 A**

**Subgroup I: 50 ADY, 51 A, 52 S // \*48 I, \*64 G**

**///// AAS, DAS, YAS (I-G for all)**

**Subgroup II: 50 TKE, 51 LVGI, 52 S // \*48 I, \*64 G**

**///// TLS, KVS, EVS, LGS, KIS (I-G for all)**

**Subgroup III: 50 GD, 51 A, 52 S // \*48 I, \*64 G**

**///// GAS, DAS (I-G for all)**

**Subgroup IV: 50 W, 51 A, 52 S // \*48 I, \*64 G**

**///// WAS (I-G)**

**Subgroup V: 50 E, 51 A, 52 T // \*48 I, \*64 G**

**///// EAT (I-G)**

**Subgroup VI: 50 Y, 51 A, 52 S // \*48 I, \*64 G**

**///// YAS (I-G)**

**L3 (91-96)**

***<< 3, 1995 >> human Vk germlines (divided into subgroups) with the information about the canonical structures of each region***

***<< 4, 7, 1995 >> lengths of L3 of each canonical structure***

***<< 4, 1995 >> key residues***

***<< 4, 1995 >> amount of sequences for each canonical structure***

**// (Jk) = position located in Jk (see below different Jk segments)**

**5 canonical structures**

1. **Canonical structure L3-1 (75% = 324)**

**// length = 6**

**// 90 QNH, 91 -, 92 -, 93 -, 94 -, 95 P, 96 -, 97 TS**

**Subgroup I: 91 SYHFALD, 92 YDN, 93 SN, 94 TLAYFT, \*95 P, 96 (Jk) // \*90 Q, \*97 (Jk)**

**///// SYSTP, YDNLP, YNSAP, HNSYP, YNSYP, FNSYP, ANSFP, LNSYP, YYSTP, YYSYP, YYSFP, DYNYP, YNSYS (+ 96 (Jk) for all) (Q for all)**

**Subgroup II: 91 RGSA, 92 ITL, 93 EHQ, 94 FWLT, \*95 P, 96 (Jk) // \*90 Q, \*97 (Jk)**

**///// RIEFP, GTHWP, GIHLP, SIQLP, ALQTP, ATQFP (+ 96 (Jk) for all) (Q for all)**

**Subgroup III: 91 YRD, 92 GNSY, 93 SN, 94 SWL, \*95 PH, 96 (Jk) // \*90 Q, \*97 (Jk)**

**///// YGSSP, YNNWP, RSNWP, RSNWH, DYNLP (+ 96 (Jk) for all) (Q for all)**

**Subgroup IV: 91 Y, 92 Y, 93 S, 94 T, \*95 P, 96 (Jk) // \*90 Q, \*97 (Jk)**

**///// YYSTP (+ 96 (Jk)) (Q)**

**Subgroup V: 91 H, 92 D, 93 N, 94 F, \*95 P, 96 (Jk) // \*90 Q, \*97 (Jk)**

**///// QHDNFP (+ 96 (Jk)) (Q)**

**Subgroup VI: 91 SG, 92 SN, 93 SK, 94 LH, \*95 P, 96 (Jk) // \*90 Q, \*97 (Jk)**

**///// SSSLP, GNKHP (+ 96 (Jk) for all) (Q for all)**

1. **Canonical structure L3-2 (0% = 1)**

**// length = 6**

**// 90 Q, 91 -, 92 -, 93 -, 94 P, 95 -, 96 -, 97 T**

1. **Canonical structure L3-3 (10% = 49)**

**// length = 5**

**// 90 Q, 91 -, 92 -, 93 -, 94 -, 95 -, 97 -**

1. **Canonical structure L3-4 (0% = 1)**

**// length = 4**

**// 90 Q, 91 -, 92 -, 93 -, 94 -, 97 S**

1. **Canonical structure L3-5 (15% = 67)**

**// length = 7**

**// 90 Q, 91 -, 92 -, 93 -, 94 -, 95 -, 96 P, 96a -, 97 T**

**Jk segments (96-108) *<< 7, 1995 >>***

1. **Jk1 WTFGQGTKVEIKR**
2. **Jk2 YTFGQGTKLEIKR**
3. **Jk3 FTFGPGTKVDIKR**
4. **Jk4 LTFGGGTKVEIKR**
5. **Jk5 ITFGQGTRLEIK**

**Al-Lazikani 1997 + Guarne 1996**

**V(k) L1 Canonical structures (26-32) *<< 3, 1997 >>***

**2, 3, 4, 6 – both human and non-human (marked with red color) // 1, 5 – only non-human *<< 8, 1997 >>***

***<< 3, 1997 >> numbering schemes***

***<< 4-5, 1997 >> canonical structures k-L1-1, k-L1-2, k-L1-3, k-L1-4***

***<< 8, 1997 >> lengths and k-L1-2 types (A, B)***

1. **k-L1-1 (length = 6)**

**// Kabat: …26, 27, 29, 30, 31, 32… (L1 24-34)**

**// Chothia\*: …26, 27, 28, 29, 30, 32…**

**SSSVSS**

1. **k-L1-2 (length = 7)**

**// Kabat: …26, 27, 28, 29, 30, 31, 32… (L1 24-34)**

**// Chothia\*: …26, 27, 28, 29, 30, 31, 32…**

**type A: ASQDISNYL, ASQDVNTAV // \*71 F**

**type B: ASQDINNYL, ASGNIHNYL // \*71 Y**

1. **k-L1-3 (length = 13)**

**// Kabat: …26, 27, 27a, 27b, 27c, 27d, 27e, 27f, 28, 29, 30, 31, 32… (L1 24-34)**

**// Chothia\*: …26, 27, 28, 29, 30, 30a, 30b, 30c, 30d, 30e, 30f, 31, 32…**

**SQSLFNSGKQKNY, SQSLLNSGNQKNF**

1. **k-L1-4 (length = 12)**

**// Kabat: …26, 27, 27a, 27b, 27c, 27d, 27e, 28, 29, 30, 31, 32… (L1 24-34)**

**// Chothia\*: …26, 27, 28, 29, 30, 30a, 30b, 30c, 30e, 30f, 31, 32…**

**SQSLVHSNGNTY, SQSIVHSSGNTY**

1. **k-L1-5 (length = 11)**

**// Kabat: …26, 27, 27a, 27b, 27c, 27d, 28, 29, 30, 31, 32… (L1 24-34)**

**// Chothia\*: …26, 27, 28, 29, 30, 30a, 30b, 30c, 30d, 31, 32…**

**no accurate high resolution structures(**

1. **k-L1-6 (length = 8)**

**// Kabat: …26, 27, 27a, 28, 29, 30, 31, 32… (L1 24-34)**

**// Chothia\*: …26, 27, 28, 29, 30, 30a, 31, 32…**

**no accurate high resolution structures(**

**V(lambda) L1 Canonical structures (25-32) *<< 8, 1997 >>***

***<< 9, 1997 >> numbering schemes***

***<< 6-7, 1997 >> lambda-L1 canonical structures***

***<< 6-7, 8, 1997 >> lengths and lambda-L1-3 types (A, B)***

1. **lambda-L1-1 (length = 10)**

**// Kabat: …25, 26, 27, 27a, 27b, 28, 29, 30, 31, 32… (L1 24-34)**

**// Chothia\*: …25, 26, 27, 28, 29, 30, 30a, 30b, 31, 32…**

**GTSSNIGSSTV, GSATDIGSNSV**

1. **lambda-L1-2 (length = 11)**

**// Kabat: …25, 26, 27, 27a, 27b, 27c, 28, 29, 30, 31, 32… (L1 24-34)**

**// Chothia\*: …25, 26, 27, 28, 29, 30, 30a, 30b, 30c, 31, 32…**

**GSSSNIGAGHNV**

1. **lambda-L1-3 (length = 11)**

**// Kabat: …25, 26, 27, 27a, 27b, 27c, 28, 29, 30, 31, 32… (L1 24-34)**

**// Chothia\*: …25, 26, 27, 28, 29, 30, 30a, 30b, 30c, 31, 32…**

**type A: SSTGAVTTSNYA**

**type B: SSTGTVTSGNHA**

1. **lambda-L1-4 (length = 9)**

**// Kabat: …25, 26, 27, 27a, 28, 29, 30, 31, 32… (L1 24-34)**

**// Chothia\*: …25, 26, 27, 30, 30a, 30b, 30c, 31, 32…**

**ANALPNQYAY**

**L2 Canonical structures (50-52) *<< 9, 1997 >>***

***<< 10, 1997 >> L2 canonical structure***

1. **L2-1 (length = 3)**

**YRDAM, YEISK, YDASN, YYTTT, YYTST, YSASF, YWAST, YKVSN, YRVSN, YGAST, YYNDL, YKDTQ, GDTNN, GGTNN**

**V(k) L3 Canonical structures (91-96) *<< 9, 1997 >>***

**1, 2, 3, 6 (4, 5 are for V(lambda) L3) *<< 1, 1996 Guarne >>***

**Information about 94-95 Pro and commentaries for cs 1, 2, 3 and 6 for Vk *<< 1, 1996 Guarne >>***

***<< 10-11, 1997 >> k-L3-1 and k-L3-2 canonical structures and their lengths***

1. **k-L3-1 (length = 6) // most commonly observed in L3 V(k)**

**// 95 P**

**QYQNLPLT, QGNTLPPT, QHYTTPPT, NDYSNPLT, QSTHVPWT, QGSHIPFT, QGTHVPYT, NDHSYPLT**

1. **k-L3-2 (length = 6)**

**// 94 P**

**QWTYPLIT**

1. **k-L3-3 (length = 5)**

**// 95 P, 5-residue loop**

**no data available(**

**see << Chothia 1989 >> and << Tomlinson 1995 >>**

1. **k-L3-6 (length = 5) *<< 1-2, 1996 Guarne >>***

**// no Pro at 94 or 95; 80% 94 L**

**QYYNLYT (there is no 95)**

**V(lambda) L3 Canonical structures (91-96) *<< 9, 1997 >>***

***<< 12, 1997 >> lambda-L3 canonical structures and their numbering scheme***

***<< 16-17, 1997 >> lengths and lambda-L3-1 types (A, B, C)***

1. **lambda-L3-1 (L3-4) (length = 6)**

**// Chothia: …91, 92, 93, 94, 95, 96…**

**type A: LWYSNLWV, LWYSNHWV**

**type B: SYDRSLRV**

**type C: LWSNNHWI**

1. **lambda-L3-2 (L3-5) (length = 8)**

**// Chothia: …91, 92, 93, 94, 95, 95a, 95b, 96…**

**AWDVSLNAYV, AWNDSLDEPG**

**H1 Canonical structures (26-32) *<< 17, 1997 >>***

**3 canonical structures *<< 17, 1997 >>***

***<< 13, 1997 >> numbering schemes and lengths***

***<< 13, 1997 >> canonical structure H1-1***

1. **H1-1 (length = 7) // the most commonly observed**

**// Kabat: …29, 30, 31, 32, 33, 34, 35, 36… (H1 31-35b)**

**// Chothia\*: …29, 30, 31, 32, 33, 34, 35, 36… (H1 26-32)**

**SGFIFSSYA, SGFDFSKYW, SGFTLSGET, SGFTFSNYG, SGFSFSSYG, SGYTFTEYT, SGFNIKDTY, SGFSLTGYG, SGYTFTTYG, SGYTFTNYW, SGFTFNIFV, SGTSFDDYY, TGYTFSEYW, SGFLLISNG, SGFTFSDYW**

1. **H1-2 (length = 8)**

**// Kabat: …29, 30, 31, 32, 33, 34, 35, 35a, 36… (H1 31-35b)**

**// Chothia\*: …29, 30, 31, 31a, 32, 33, 34, 35, 36… (H1 26-32)**

**no accurate high resolution structures(**

1. **H1-3 (length = 9)**

**// Kabat: …29, 30, 31, 32, 33, 34, 35, 35a, 35b, 36… (H1 31-35b)**

**// Chothia\*: …29, 30, 31, 31a, 31b, 32, 33, 34, 35, 36… (H1 26-32)**

**no accurate high resolution structures(**

**H2 Canonical structures (52-56) *<< 17, 1997 >>***

**4 canonical structures *<< 17, 1997 >>***

***<< 14-15, 1997 >> H2 canonical structures and their numbering schemes***

***<< 17, 1997 >> lengths and H2-2 types (A, B), H2-3 types (A, B, C)***

1. **H2-1 (length = 5)**

**// Chothia: …52, 53, 54, 55, 56…**

**FYTGT, WGDGN, WAGGN**

1. **H2-2 (length = 6)**

**// Chothia: …52, 52a, 53, 54, 55, 56…**

**type A: YPGNS, YPTNG, NPKNG,NTYSG (… - 56)**

**type B: LPGSG (… - 56)**

1. **H2-3 (length = 6)**

**// Chothia: …52, 52a, 53, 54, 55, 56…**

**type A: WDDGSD, WYNGSR, HPDSGT, SNGGGY**

**type B: LSGGGF**

**type C: FGSGGN**

1. **H2-4 (length = 8)**

**// Chothia: …52, 52a, 52b, 52c, 53, 54, 55, 56…**

**RNKPYNYE**

**\*In the original article the second numbering scheme is called “Structure”, i.e. structurally correct = therefore, this meant to be the Chothia numbering scheme**

**H3 Canonical structures (95-102? // 96-101) *<< 18, 1997 >>***

**“**The relations between the sequences and structures of the H3 hypervariable region are not as well understood as those for the other hypervariable regions.**” *<< 18, 1997 >>***

**Extra R**

**invariant residues: 23/22 C, 35/36 W, 38/39 Q, 82/86 D, 86/90 Y, 88/92 C, 99/104 G, 101/106 G, 102/107 T**

**closely related residues: 4/4 LM, 6/6 QE, 16/15 GS, 21/20 VLIM, 22/21 ST, 47/48 VLIM, 61/66 RK, 63/68 ST, 65/70 ST, 73/80 LIMF, 75/82 LIM, 78/82c VLM, 84/88 GA, 87/91 YF, 98/103 FW, 104/109 VL, 106/111 VLI**

***bn* sites (C, V ,L, I, M, F, W + P, H ,Y ,G, A, S, T) = 11/10 (+RKE), 12/11, 13/12 (+RKE), 15/14 (P), 19/18 (VLIM), 25/24, 33/34, 36/37, 44/45, 46/47, 48(I)/49(GAS), 62/67 (*b*), 64(G)/69(MI), 71(A if (lambda), FY if (k))/78, 83/87, 85/89, 97/102**

***n* sites (P, H, Y ,G, A, S, T) = 40/41 (PS), 72(STA)/79(YF)**

***sn* sites (R, K, E, D, Q, N + P, H ,Y ,G, A, S, T) = 5/5, 14/13, 17/16, 18/17, 20/19, 24/23, 34/35, 39/40, 41/42 (G), 42/43, 43/44 (*n*), 60/65, 66/71, 67(S)/72(D), 68/73, 69/74, 70/77, 74/81, 76/82a, 77/82b, 79/83, 80/84, 81/85, 89/93, 90/94, 100/105, 103/108, 105/110, 107/112**

***s* sites (R, K, E, D, Q, N) = 37/38 (RKQ), 45/46**

**(s)-(n)-(b) groups:**

**(s) surface group: R, K, E, D, Q, N (high probability to be on the surface)**

**(n) neutral group: P, H, Y, G, A, S, T (equal probabilities)**

**(b) buried group: C, V, L, I, M, F, W (high probability to be buried)**

**Afterword: m/n – light/heavy** chain in the Chothia numbering ///// y(X) means that X (or X-group) is present at y residue rather than other possible ((if there a space before (X), it means that it concerns to both light and heavy chains))

*<< 5, 1998All >> s, n, b groups of amino acids*

*<< 11, 1998All >> invariant and closely related residues in VH, VL*

*<< 11, 1998All >> positions of bn, n, sn, s sites*

*<< 10, 12, 15, 17, 19, 21, 1998All >> more info about some positions, i.e.*

*Rather than amino acids of their groups:*

*10 - 40/41 P, 43/44 n, 62/67 b*

*12 - 67Light S, 72Light STA, 72Heavy D, 79Heavy YF*

*15 - 19/18 VLIM, 37/38 s (usually RKQ)*

*17 - 71/78 A in V(lambda) and FY in V(k)*

*19 - 15/14 P, 16/15 GS, 40/41 PS, 41/42 G*

*21 - 48Light I, 64Light G; 49Heavy GAS, 69Heavy MI*

***//add this info to other positions later << 6-9, 1998All >>***

**Extra RR (Chothia specified)\***

**L1** – **G**, DEQ, RKQTS, VA, TS, ILM, STN, **C (23)**, RKST, ASG = **S**T … YNFAW = LVMIA, ANH, **W (35)**, **Y**FLV, QL, **Q**EH, **K**R, **P**SQ, **G**DH, QKGT */// for V(lambda) L1 starts one position earlier*

**L2** – **G**DH, QKGT, SAPT, **P**FY, KRQT, LRGTV, **L**WV, **I**VM, **Y**KG (49) = YKW, ATV … LRS (54), AF, STDP, **G (57)**, **V**I, **P**, DSAV, **R**, **F**, **S**T, **G**

**L3** – **E**D, **D**, LFEIAV, AG, TVDI, **Y**, YF, **C (88)**, QFLAS, QHL = YGSWH … **P**LH, YLPRWF = **T**V, **F (98)**, **G**, GQAST, **G**, **T**, **K**R, **L**V, **E**TD

**H1** – **G**S, GAQ, **S**T, LVM, KRS, LIMV, **S**T, **C (22)**, KATS, AVT, **S**T (25) = … = YWGA (33), MIWV, HNSG, **W (36)**, **V**IF, RK, **Q**K, ARPFST, **P**H, **G**E, KQN

**H2** – **G**E, KQNH, GRKEA, **L**R, **E**K, **W**YG, IVML, GA, YW, **I**V (51) = SNYD … SNYDT = TIPSKA (57), YN, **Y**F, NASVG, DPEQA, SKADT, FVL, **K**QR, GSD (65), RK

**H3** – **E**DA, **D**, TS, **A**G, VTIML, **Y**, **Y**F, **C (92)**, ATV, RASN = GW … FMGLY, DAGV, YV = **W (103)**, **G**, **Q**AEKHP, **G**, **T**, TLSQ, **V**L, **T**

**Afterword: X** – CDR region and its neighborhood - each residue position is separated with other by “,” ///// for each position, the most common amino acids are listed, **but!** the variability of the amino acids doesn’t completely correlate between different positions

///// there are three types of marking here: a) **black and just bolded** – the possibility of this residue at this position is high b) **purple and bolded** - the possibility of this residue at this position is super high (mostly invariant) c) **purple, bolded and bigger** - the possibility of this residue at this position is also super high, but it’s almost noticed as invariant referring to the original rules (Kabat, Chothia and IMGT)

///// **the region marked with a red color corresponds to the CDR (=\*)**

///// some positions are numbered (in the Kabat/Chothia numberings (at these positions they are identical)) to help with finding these residues in the sequences (other positions can be numbered just with using the sequence of natural numbers (but without crossing the ellipsis bound)**\*\***)

**\*numbering and CDRs boundaries considered in this article are different from the Chothia ones, so the obtained info was accurately translated into the Chothia annotation**

**\*\*it can be done, because regions with structural insertions (deletions) those make all the differences in the numberings, are absent here or located in the region replaced with ellipsis**

**This information can be only an addition to the main rules!**

*<< 3, North >> Extra RR*

*<< 26-27, North >> Extra RR additions (i.e. more possible amino acids for some of considered positions)*

**Extra RRR (Chothia specified)\***

**Clusters for CDRs \*\***

**L1** – length 6-13 (mostly 7 or 12)

**///// 24-34 ([24-25] FR1 + L1 + [33-34] FR2 for V(k))**

**///// 24-34 ([24] FR1 + L1 + [33-34] FR2 for V(lambda))**

(L1-2A) L1-11-1 – R**ASQ**DISNYLA (76, k, HM) ///// F71

(L1-4) L1-16-1 – **RSS**QSLVHSN**G**N**TYL**E (68, k, HM)

(L1-2B) L1-11-2 – R**AS**QD**I**SNY**L**N (55, k, M) ///// T/G 71

(L1-3) L1-17-1 – **KSSQSL**LN**S**RTRK**NYLA** (21, k, HM)

(L1-1) L1-10-1 – S**A**S**SSV**S**Y**MH (20, k, M)

(L1-3λA, L1-3λB) L1-14-1 – S**A**S**SSV**S**Y**MH (14, lambda, H)

(L1-5) L1-15-1 – R**AS**E**SVD**SY**G**N**S**F**M**N (11, k, HM)

(L1-1λ) L1-13-1 – **SG**SS**SNIG**N**N**Y**V**S (7, lambda, H)

L1-12-1 – R**AS**S**S**V**SS**SYLH (5, k, M)

L1-12-2 – R**AS**Q**S**VSSNYL**A** (5, k, HM)

(L1-4λ) L1-11-3 – SGNNLGS-SVH (5, lambda, H)

L1-13-2 – **TRSSG**N**I**AS**NYV**Q (4, lambda, H)

(L1-2λ) L1-14-2 – S**A**S**SSV**S**Y**MH (4, lambda, M)

L1-10-2 – **SASSSVSY**MY (2, k, M)

(L1-6) L1-12-3 – **TLS**S**QHSTYTIE** (2, lambda, HM)

L1-15-2 – **RASKSVSTSGY**N**YMH** (2, k, M)

*<< 7, North >> L1’s clusters*

*<< 8, 12-13, North >> commentaries for some L1 clusters*

***//add more detailed commentaries later***

**L2** – length only 3 (or 7?)

**///// 49-56 ([49] FR2 + L2 + [53-56] FR3)**

(L2-1) L2-8-1 (mostly this) – **Y**-ASNLAS (290, k, HM)

(L2-1) L2-8-2 – YAASNLDS (9, k, HM)

L2-8-3 – SEG**N**TLR**P** (3, k/lambda, M)

L2-8-4 – G**G**TN**NR**VP (2, k/lambda, M)

L2-8-5 – **Y**SA**S**Y**R**Y**S** (2, k, HM)

L2-12-2 – ELKKDGSHSTGD (2, lambda, M)

*<< 8, North >> L2’s clusters*

*<< 13, North >>* ***//add more detailed commentaries later***

**L3** – length 4-10 (mostly 6)

**///// 89-97 ([89-90] FR3 + L3 + [97] FR4)**

(L3-1) L3-9-cis7-1 (mostly this) – Q**Q**GSS-**P**L**T** (219, k, HM)

(L3-1λA, L3-1λB, L3-1λC) L3-9-1 – ALW-SNHWV (22, k/lambda, HM)

L3-8-1 – L**Q**YYNLR**T** (15, k, HM)

L3-9-2 – Q**Q**STH-PP**T** (12, k, HM)

(L3-2λ) L3-11-1 – AAWDSSLDAVV (9, lambda, H)

(L3-1) L3-9-cis7-2 – **QH**FWS**TP**R**T** (8, k, HM)

L3-10-1 – QSYDSS-SVV (6, lambda, H)

(L3-3) L3-8-cis6-1 – **Q**QWNY**P**F**T** (3, k, M)

L3-13-1 – AAW**D**DSRGGPDW**V** (3, lambda, HM)

(L3-4) L3-7-1 – Q**Q**YN**SY**S (2, k, HM)

L3-9-cis7-3 – Q**Q**YYIY**P**Y**T** (2, k, HM)

(L3-5) L3-10-cis8-1 – LYSREF**PP**W**T** (2, k, M)

(L3-2) L3-9-cis6-1 – **QQWTYPLIT** (1, k, M)

L3-10-cis7,8-1 – **SQSTHVPPLT** (1, k, M)

L3-11-cis7-1 – **QQYNNWPPRYT** (1, k, H)

L3-12-1 – **ATWDSGLSADWV** (1, lambda, H)

*<< 9, North >> L3’s clusters*

*<< 13-14, North >>* ***//add more detailed commentaries later***

**H1** – length 4-10 (mostly 7)

**///// 23-35 ([23-25] FR1 + H1 + [33-35] FR2)**

(H1-1) H1-13-1 (mostly this) – KA**SG**FTFTDYYMH (267, HM)

(H1-2) H1-14-1 – **TVTGYSIT**SG**Y**A**W**N (11, M)

(H1-3) H1-15-1 – SF**SGFS**LSTSGMG**V**G (9, HM)

(H1-1) H1-13-2 – KA**S**GFNITDYYIS (7, HM)

H1-13-3 – KA**SG**YT**F**TTYAMN (5, HM)

H1-13-4 – AVS**G**FSFSGYYWS (4, HM)

H1-13-5 – A**ASG**FTYSINYMG (4, HM)

H1-13-6 – **A**A**SG**YKYTNYCM**G** (4, C)

H1-13-7 – SVT**G**DSI**TS**GYWN (3, M)

H1-13-8 – KA**SG**YTFTTYDMG (3, M)

H1-13-9 – **A**A**SG**N**T**LSTYDMG (3, CL)

H1-13-10 – **KASGGTFS**M**Y**GFN (2, H)

H1-13-11 – K**AS**EY**T**LTSYLFQ (2, M)

H1-13-cis9-1 – **AASGYTIGPYCMG** (2, C)

H1-10-1 – A**AS**T**YT**DTV**G** (2, C)

H1-12-1 – **KLWYTFTDYGMN** (1, M)

H1-16-1 – **AASGRAASGHGHYGMG** (1, L)

*<< 10, North >> H1’s clusters*

*<< 14, North >>* ***//add more detailed commentaries later***

**H2** – length 5-12 (mostly 7 and 6)

**///// 51-58 ([51] FR2 + H2 + [57-58] FR3)**

(H2-2A) H2-10-1 – -**I**YPGNG-T- (155, HM) ///// AVLISTQ 71 (mostly)

(H2-1) H2-9-1 – Y**I**WYS**G**STY (77, HM)

(H2-3A, H2-3C) H2-10-2 – -**I**SSGGGNTY (42, HM) ///// **R**D 71 (mostly)

(H2-4) H2-12-1 – E**IR**N**K**ANNYT**T**E (26, M)

(H2-2B) H2-10-3 – E**I**L**PG**SGSTN (11, HM)

H2-10-4 – T**I**SSG**G**GYTN (7, M)

H2-10-5 – **A**ISG**GG**TYIH (3, MC)

H2-10-6 – RIDPN**G**GG**TK** (3, HM)

(H2-3B) H2-10-7 – **T**TLS**G**GGF**T**F (2, HM)

H2-10-8 – G**I**D**P**HN**GG**GA (2, HM)

H2-10-9 – G**I**DPHNGGPV (2, HM)

H2-8-1 – TILG**GS**TY (2, H)

H2-9-2 – S**I**YNGFRIH (2, M)

H2-9-3 – Y**I**RYG**G**GT**Y** (2, MC)

H2-15-1 – **TIGRNLVGPSDFYTR** (1, L)

*<< 11, North >> H2’s clusters*

*<< 14-16, North >> commentaries for some H2 clusters*

***//add more detailed commentaries later***

**H3** – length 3-24 (mostly 5-14)

**///// 93-102 ([93-94] FR3 + H3)**

H3-anchor-1 (mostly this) – A**R**- … YFDY (204) ///// bulged (e.g. mostly when K/R 94 and D101?)

H3-anchor-2 – ARY … DFD**Y** (35) /////(non-bulged)

H3-anchor-3 – ARG … YFDY (25)

H3-anchor-4 – ANW … DG**D**Y (24)

H3-anchor-5 – VR- … -RDY (12)

H3-anchor-6 – AS- … SFAY (6)

H3-anchor-7 – **AR**R … GFDY (4)

H3-anchor-cis4-1 – **AR**E … **P**F**D**Y (2)

(Anchor = first 3 and last 4 residues of H3)

*<< 12, North >> H3’s anchors*

*<< 16-17, North >> H3’s length and commentaries for some H3 anchors*

***//add more detailed commentaries later***

**Afterword:** for each CDR you can see several amino acids sequences that usually appear there (in the CDR), e.g. the different CDR’s clusters (idea similar to the canonical structures concept)

///// (name of Chothia canonical structure if exists) cluster’s name (according to this article) – XYZ…RG (amino acids sequence, CDRs are marked with a red color**\***; the residue written **in bold** means that within the framework of the cluster, it’s usually invariant(>90%) // then (x, k/lambda, HMCL): x – the number of proteins used in the work**\*** and having CDRs similar to the considered cluster (it’s useful to compare the occurrences of the clusters (of the considered CDR)) // k/lambda – type of the light chain (if it is) //the organisms-owners of the proteins used in the work(**\***) – H = Human, M = Mouse, C = Camel, L = Llama // + (///// some commentaries)

///// lengths that are written nearly the CDR’s name are just above the truth; their real sizes corresponding to the Kabat numbering can be found above, in the main rules

///// don’t forget that the CDR of the real sequence can differ from its expected cluster a lot // also it’s possible (and happens frequently) that none of these clusters suits the sequence’s CDR

Therefore, use these clusters **only** to find the hypothetical CDR or to prove the already found one (and in no case for a refutation)!

**\*North, B. et al. (2011). A New Clustering of Antibody CDR Loop Conformations.**

**\*CDRs are marked according to the Chothia numbering**

**\*\*it’s important that the canonical structures and clusters for CDRs are just an approximation – so they can be used only to find the hypothetical CDR or to prove the already found one**