**Chothia Rules**

**+Extra R (Chothia specified) /1998**

**+Extra RR (Chothia specified) /2011**

**+ Extra RRR – the newest canonical structures /2011**

**+ Kabat-Chothia numbering (with changed 31abc only)**

**Chothia 1987**

**// information usually based only on several structures (but discussed using thoughts about proteins’ structure) or based on statistics received from Kabat’s database**

26 - 32, 50 – 52, 91 - 96 – L1, L2, L3

26 - 32, 53 – 55, 96 - 101 – H1, H2, H3

*<< 4, 1987 >>*

***L1***

**V(lambda)**

*//only 3 structures*

**25 G**, 26 ST, 27 AS, 28 TS, 29 DNS, 30 IV, 30a G, 30b SA, 30c -G, 31 NI, 32 STH, **33 VI // 71A**

*<< 5, 1987 >>***; Ile 30** *<< 8, 1987 >>*

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

**V(k)**

*//only 3 structures*

25 AS, 26 S, 27 SE, 28 SD, 29 VIL, 30 SIL, 31 -KN, 31a -S, 31b -G, 31c -N, 31d -E, 31e -K, 31f -N, 32 SYF, 33 L(V) // 2 I(MV), 71 F(Y) *<< 7, 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 >>*

**L1 (k) length: 7, 8, (11, 12, 13)**

**L1 (k) length mouse: 6, 7, 11, 12, 13** *<< 7, 1987 >>*

***L2***

**(50 YREG, 51 NDAI, 52 DAS) //inf only from 4 structures! // 48 I, 64 GA**

***<< 7, 1987 >> // << 8, 1987 >>***

***L3***

**V(lambda)**

***//only 3 structures***

**91 YW, 92 DN, 93 RSD, 93a –S, 93b –DL, 94 SGND, 95 LSE, 96 RYP *<< 8, 1987 >>***

***<< 9, 1987 >>* - 94 GN**

**L3 length: 6, 7, 8, (5)**

**L3 length mouse: 6 *<< 9, 1987 >>***

**V(k)**

***//only 3 structures***

**90 QN, 91 YDW, 92 QHT, 93 SY, 94 LYP, 95 PL, 96 *YLI << 9, 1987 >>***

**L3 length: 6, (5)**

**L3 length mouse: 6, (5, 7, 12) *<< 10, 1987 >>***

***//based on Kabat’s statistics***

**90 QN, 95 P *<< 10, 1987 >>***

***H1***

***//only 4 structures***

**26 G, 27 FT, 28 ITDS, 29 F, 30 SD, 31 SDK, 32 YF // 34 M(I), 94 R(K*)* *<< 10, 1987 >>***

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

***H2***

**H2 length: 4, 6 (all mouse’s), 3, (5) *<< 10, 1987 >>***

***//only several sructures or based on Kabat’s statistics***

**50 Y, 51 V, 52 FHW, 52a PD, 53 YD, 54 HSG, 55 GNDS, 56 TD, 57 S, 58 *D* *<< 10-11, 1987 >>***

***H3***

***HUMAN (JH segments)***

JH1: AEYFGHWGQGTLVTVSS

JH2: YWYFDLWGRGTLVTVSS

JH3: AFDVWGQGTMVTVSS

JH4: YFDYWGQGTLVTVSS

JH5: NWFDSWGQGTLVTVSS

JH6: YGMDVWGQGTTVTVSS ***<< 12, 1987 >>***

**MOUSE *(JH segments)***

J(H1): WYFDVWGAGTTVTVSS

J(H2): YFDVWGQGTTVTVSS

J(H3): WFAYWGWGTLVTVSA

J(H4): DYWGWGTSVTVSS ***<< 12, 1987 >>***

***//only 4 structures***

**length: 6, 7, 9, 15 *<< 12, 1987 >>***

***//only several structures but were discussed using thoughts about protein’s structure interactions***

**94 R(K), 95 N, 96 Y, 97 Y, 98 G, 99 S, 100 T, 100a W, 100b Y(FW), 100c FL(M), 101 D, 102 V**

**// may be: 101-1 M, 101-2 FW *<< 12, 1987 >> + << 13, 1987 >>***

**H3 length: 10, 6, 9 7, (8, 11, 12, 14) *<< 13, 1987 >>***

**Extra *<< 6, 1987 >> - residues commonly buried within VL and VH domains***

**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**

**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 (52-55) (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)**

**H1 (26-32) *<< 3, 1992 >>***

**H2 (52-56) *<< 5, 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)**

**\* - key residues (residues at key sites)**

**//to get the other (not human) canonical structure classes and families, see refs on *<< 3, 1992 >>***

**Chothia (Tomlinson 1995)**

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

**L1 (26-32)**

***<< 3, 1992 >> - human Vk germlines (divided into subgroups) with the information about the canonical structures of each region***

***<< 3, 1992 >> - lengths of L1 of each canonical structure***

***<< 4, 1992 >> - 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, 1992 >> - human Vk germlines (divided into subgroups) with the information about the canonical structures of each region***

***<< 4, 6, 1992 >> - length of the L2 canonical structure***

***<< 4, 1992 >> - 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, 1992 >> - human Vk germlines (divided into subgroups) with the information about the canonical structures of each region***

***<< 4, 7, 1992 >> - lengths of L1 of each canonical structure***

***<< 4, 1992 >> - key residues***

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

**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**

**Chothia (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 >>***

***<< 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)**

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

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

**QWTYPLIT**

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

**no data available(**

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

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

**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**

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

**see << 1996 Shirai >> and << 1997 Morea >> // + (not Chothia’s annotation method) << 1996 Martin and Thornton >>**

**Chothia (1998 Morea) << check if it useless later>>**

**H3 – between 92Cys and 104Gly (here, extended H3 – 92-104) *<< 1, Morea >>***

**94 – Arg/Lys, 101 – Asp *<< 1, Morea >>***

**4 residues from the N terminus and 6 from the C terminus – torso of the region (e.g. we know smth about it) *<< 5, Morea >>***

**The Bulged Torso Conformation**

**Arg94, Asp101,Trp103 +exception (93Lys and 101Asp) *<< 5, Morea >>***

**if Asp101(or absent) and Arg/Lys 94 – bulged**

**if Arg/Lys 94 absent – non-bulged**

**(but there are exceptions)**



***<< 5, Morea >>***

**94 – RK, 101 – D *<< 5, Morea >>***

**Light chain: 49 – Y (usually), 99 – F, 103 – W (in all known Fab structures) ///// 97 – S, D *<< 11, Morea >>***

**100b – Y, W (or large hydrophobic, G(?)), Ser (exception?) *<< 11, Morea >>***

**Non-Bulged Torso Conformation**

**… *<< 12, Morea >>***

**//about structure and functions but without naming residues *<< …-23, Morea >>***