Supplementary Information
for

Structural Conservation of Insulin/IGF Signalling Axis at the Insulin Receptors Level in *Drosophila* and Humans

Cristina M. Viola[1,5], Orsolya Frittmann[1,2], Huw Jenkins[1],Talha Shafi[1], Pierre De Meyts[3,4], Andrzej M. Brzozowski[1*]

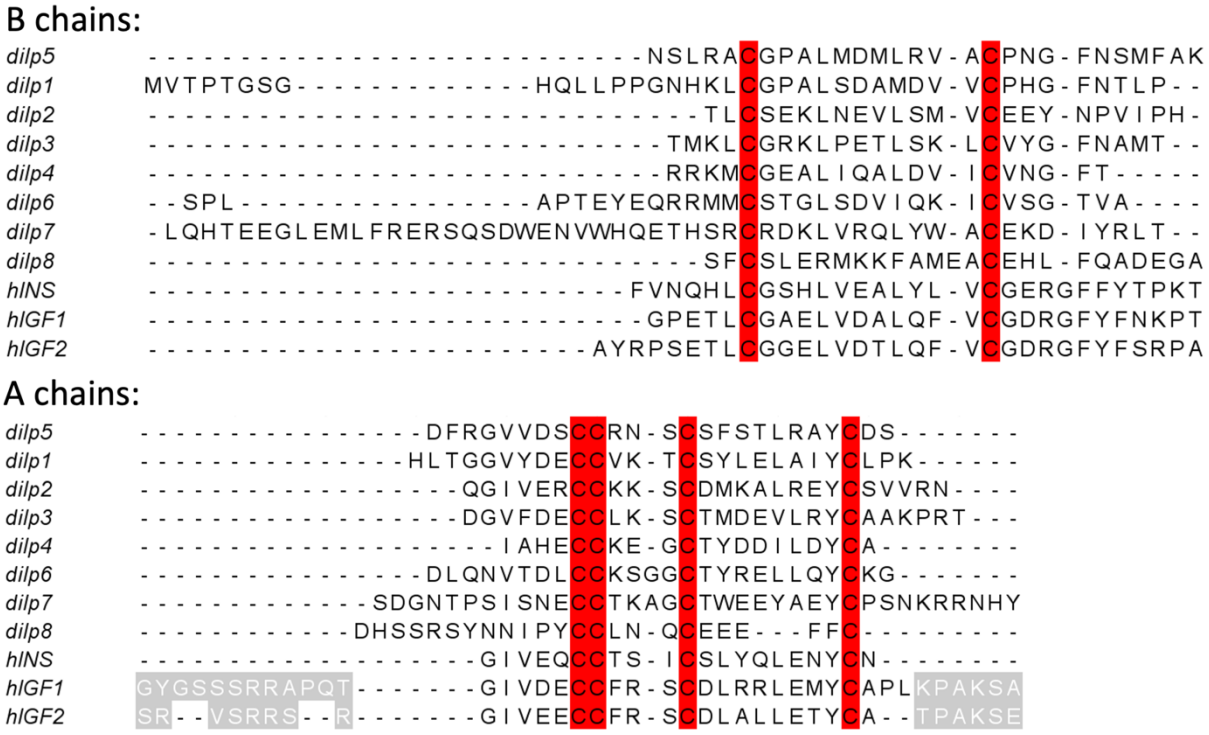
[1] York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5DD, United Kingdom

[2] Current address: Department of Haematology, University Medical Center Groningen, , Hanzeplein 1, 9713 GZ Groningen, Netherlands

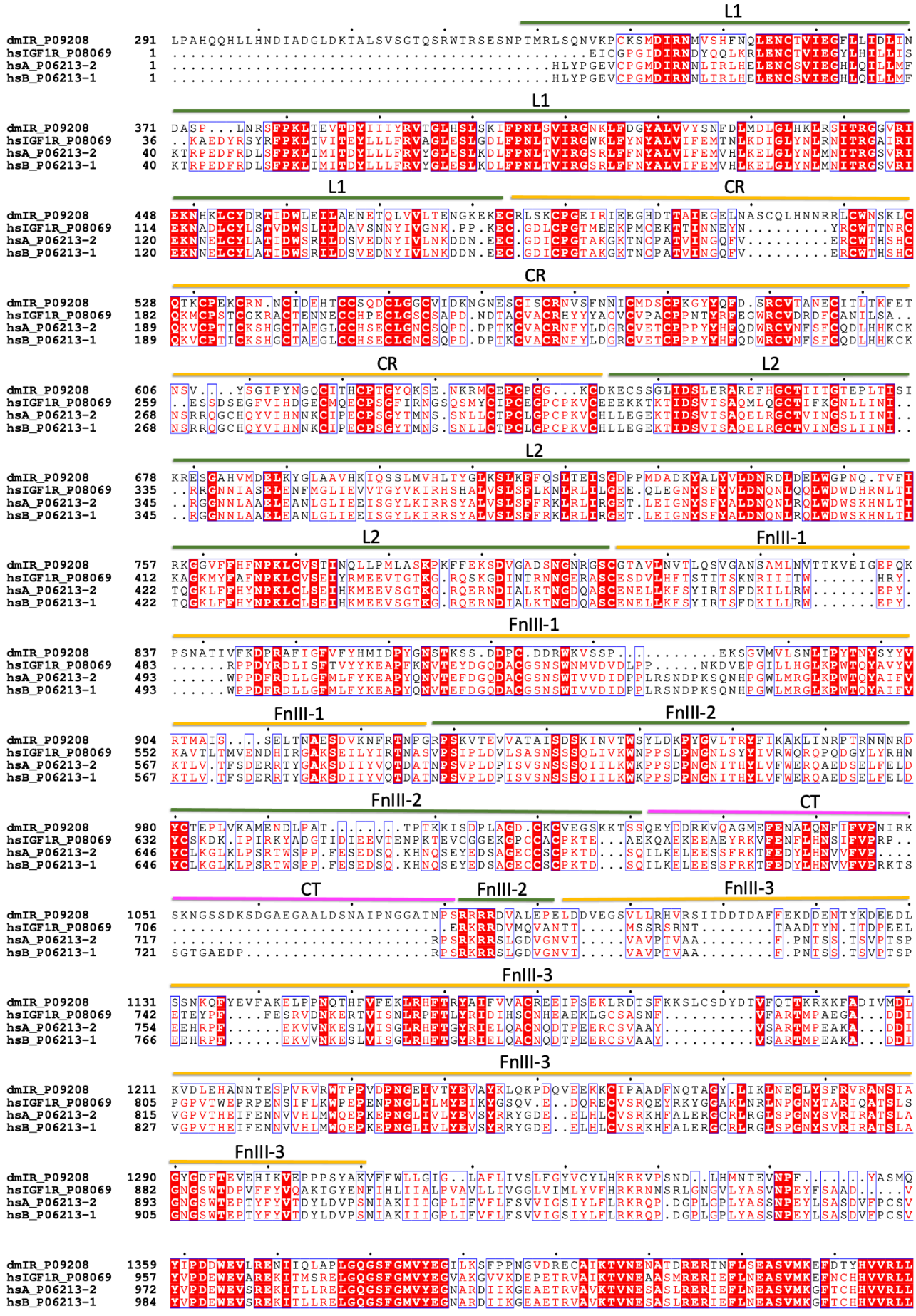
[3] Department of Cell Signalling, de Duve Institute, B-1200 Brusssels, Belgium.

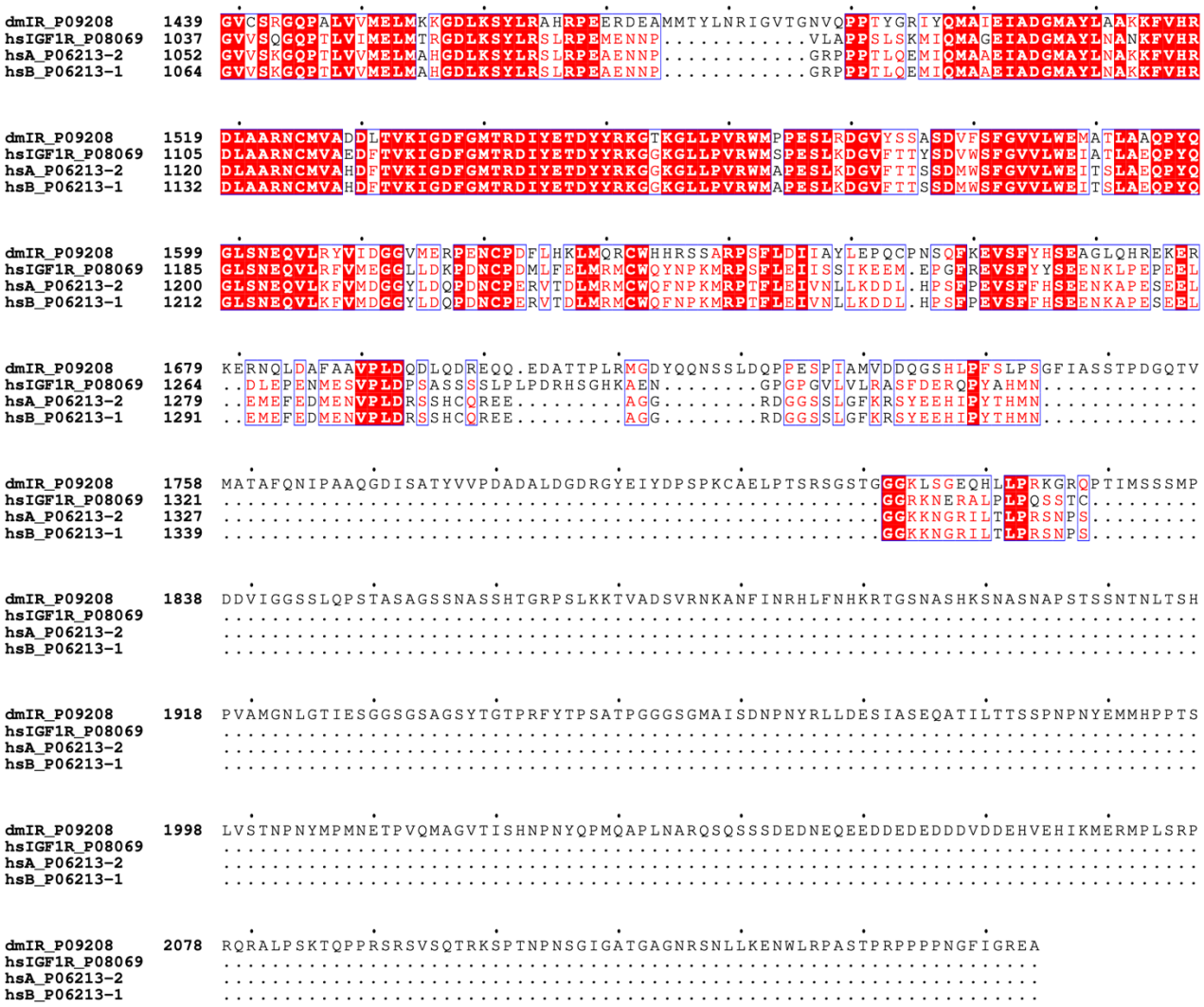
[4] Department of Cell Therapy Research, Novo Nordisk A/S, DK-2670 Maaloev, Denmark.

[5] Current address: Department of Biology, University of York, Heslington, York YO10 5DD, United Kingdom



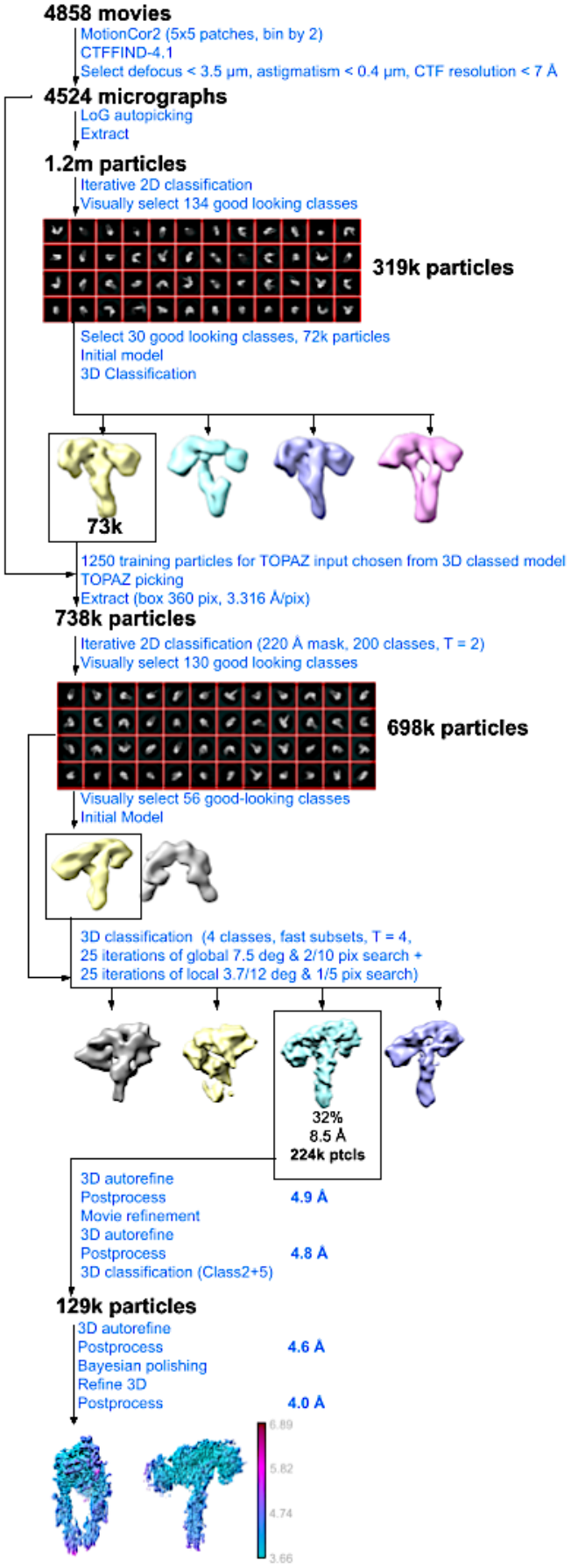
SI Figure 1. Sequence alignment of DILP1-8, human insulin (hINS) and human IGF1 (hIGF1) and IGF2 (hIGF2). C- and D-domains of IGFs are in gray (in A-chains panel).


L1
dmIR_P09208 291 LPAHQQHLLHNDIADGLDKTALSVSGTQSRWTRSESNPTMRLSQNVKPCKSMDIRNMVSHFNQLENCTVIEGFLLIDLIN
hsIGF1R_P08069 1 .....EICGPGIDIRNDYQQLKRLENCTVIEGYLHILLIS
hsA_P06213-2 1 .....HLYPGEVCPGMDIRNNLTRLHELENCSVIEGHLQILLMF
hsB_P06213-1 1 .....HLYPGEVCPGMDIRNNLTRLHELENCSVIEGHLQILLMF
L1
dmIR_P09208 371 DASP...LNRSFPKLTEVTDYIIIYRVTGLHSLSKIFPNLSVIRGNKLFDGYALVVYSNFDLMDLGLHKLRSITRGGVRI
hsIGF1R_P08069 36 ..KAEDYRSYRFPKLTVITEYLLLFRVAGLESLGDLFPNLTVIRGWKLFYNYALVIFEMTNLKDIGLYNLRNITRGAIRI
hsA_P06213-2 40 KTRPEDFRDLSFPKLIMITDYLLLFRVYGLESLKDLFPNLTVIRGSRLFFNYALVIFEMVHLKELGLYNLMNITRGSVRI
hsB_P06213-1 40 KTRPEDFRDLSFPKLIMITDYLLLFRVYGLESLKDLFPNLTVIRGSRLFFNYALVIFEMVHLKELGLYNLMNITRGSVRI
L1
CR
dmIR_P09208 448 EKNHKLCYDRTIDWLEILAENETQLVVLTENGKEKECRLSKCPGEIRIEEGHDTTAIEGELNASCQLHNNRRLCWNSKLC
hsIGF1R_P08069 114 EKNADLCYLSTVDWSLILDAVSNNYIVGNK.PP.KECGDLCPGTMEEKPMCEKTTINNEYN.........YRCWTTNRC
hsA_P06213-2 120 EKNNELCYLATIDWSRILDSVEDNYIVLNKDDN.EECGDICPGTAKGKTNCPATVINGQFV.........ERCWTHSHC
hsB_P06213-1 120 EKNNELCYLATIDWSRILDSVEDNYIVLNKDDN.EECGDICPGTAKGKTNCPATVINGQFV.........ERCWTHSHC
CR
dmIR_P09208 528 QTKCPEKCRNNCIDEHTCCSQDCLGGCVIDKNGNESCISCRNVSFNNICMDSCPKGYYQFD.SRCVTANECITLTKFET
hsIGF1R_P08069 182 QKMCPSTCGKRACTENNECCHPECLGSCSAPD.NDTACVACRHYYYAGVCVPACPPNTYRFEGWRCVDRDFCANILSA..
hsA_P06213-2 189 QKVCPTICKSHGCTAEGLCCHSECLGNCSQPD.DPTKCVACRNFYLDGRCVETCPPPYYHFQDWRCVNFSFCQDLHHKCK
hsB_P06213-1 189 QKVCPTICKSHGCTAEGLCCHSECLGNCSQPD.DPTKCVACRNFYLDGRCVETCPPPYYHFQDWRCVNFSFCQDLHHKCK
CR
L2
dmIR_P09208 606 NSV..|..YSGIPYNGQCITHCPTGYQKSE.NKRMCEPCPGG...KCDKECSSGLIDSLERAREFHGCTIITGTEPLTISI
hsIGF1R_P08069 259 ..ESSDSEGFVIHDGECMQECPSGFIRNGSQSMYCIPCEGPCPKVCEEEKKTKTIDSVTSAQMLQGCTIFKGNLLINI..
hsA_P06213-2 268 NSRRQGCHQYVIHNNKCIPECPSGYTMNS.SNLLCTPCLGPCPKVCHLLEGEKTIDSVTSAQELRGCTVINGSLIINI..
hsB_P06213-1 268 NSRRQGCHQYVIHNNKCIPECPSGYTMNS.SNLLCTPCLGPCPKVCHLLEGEKTIDSVTSAQELRGCTVINGSLIINI..
L2
dmIR_P09208 678 KRESGAHVMDELKYGLAAVHKIQSSLMVHLTYGLKSIKFFQSLTEISGDPPMDADKYALYVLDNRDLDELWGPNQ.TVFI
hsIGF1R_P08069 335 ..RRGNNIASELENFMGLIEVVTGYVKIRHSHALVSLSFLKNLRLILGEE.QLEGNYSFYVLDNQNLQQLWDWDHRNLTI
hsA_P06213-2 345 ..RGGNNLAAELEANLGLIEEISGYLKIRRSYALVSLSFFRKLRLIRGET.LEIGNYSFYALDNQNLRQLWDWSKHNLTI
hsB_P06213-1 345 ..RGGNNLAAELEANLGLIEEISGYLKIRRSYALVSLSFFRKLRLIRGET.LEIGNYSFYALDNQNLRQLWDWSKHNLTI
L2
FnIII-1
dmIR_P09208 757 RKGGVFFHFNPKLCVSTINQLLPMLASKPKFFEKSDVGADSNGNRGSCGTAVLNVTLQSVGANSAMLNVTTKVEIGEPQK
hsIGF1R_P08069 412 KAGKMYFAFNPKLCVSEIYRMEEVTGTKG.RQSKGDINTRNNGERASCESDVLHFTSTTTSKNRIIITW.......HRY.
hsA_P06213-2 422 TQGKLFFHYNPKLCLSEIHKMEEVSGTKG.RQERNDIALKTNGDQASCENEILKFSYIRTSFDKILLRW.......EPY.
hsB_P06213-1 422 TQGKLFFHYNPKLCLSEIHKMEEVSGTKG.RQERNDIALKTNGDQASCENELLKFSYIRTSFDKILLRW.......EPY.
FnIII-1
dmIR_P09208 837 PSNATIVFKDPRAFIGFVFYHMIDPYGNSTKSS.DDPC.DDRWKVSSP.....EKSGVMVLSNLIPYTNYSYYV
hsIGF1R_P08069 483 ......RPPDYRDLISFTVYYKEAPFKNVTEYDGQDACGSNSWNMVDVDLPP.....NKDVEPGILLHGLKPWTQYAVYV
hsA_P06213-2 493 ......WPPDFRDLLGFMLFYKEAPYQNVTEFDGQDACGSNSWTVVDIDPPLRSNDPKSQNHPGWLMRGLKPWTQYAIFV
hsB_P06213-1 493 ......WPPDFRDLLGFMLFYKEAPYQNVTEFDGQDACGSNSWTVVDIDPPLRSNDPKSQNHPGWLMRGLKPWTQYAIFV
FnIII-1
FnIII-2
dmIR_P09208 904 RTMAIS...SELTNAESDVKNFRTNPGRPSKVTEVVATAISDSKINVTWSYLDKPYGVLTRYFIKAKLINRPTRNNNRD
hsIGF1R_P08069 552 KAVTLTMVENDHIRGAKSEILYIRTNASVPSIPLDVLSASNSSSQLIVKWNPPSLPNGNLSYYIVRWQRQPQDGYLYRHN
hsA_P06213-2 567 KTLV.TFSDERRTYGAKSDIIYVQTDATNPSVPLDPISVSNSSSQIILKWKPPSDPNGNITHYLVFWERQAEDSELFELD
hsB_P06213-1 567 KTLV.TFSDERRTYGAKSDIIYVQTDATNPSVPLDPISVSNSSSQIILKWKPPSDPNGNITHYLVFWERQAEDSELFELD
FnIII-2
CT
dmIR_P09208 980 YCTEPLVKAMENDLPAT.|..|..|.TPTKKISDPLACD.CKCVEGSKKTSSQEYDDRKVQAGMEFENALQNFIFVPNIRK
hsIGF1R_P08069 632 YCSKDK.IPIRKYADGTIDIEEVTENPKTEVCGGEKGPCCACPKTE...AEKQAEKEEAEYRKVFENFLHNSIFVPRP..
hsA_P06213-2 646 YCLKGLKLPSRTWSPP.FESEDSQ.KHNQSEYEDSACECCSCPKTD...SQILKELEESSFRKTFEDYLHNVVFVP....
hsB_P06213-1 646 YCLKGLKLPSRTWSPP.FESEDSQ.KHNQSEYEDSACECCSCPKTD...SQILKELEESSFRKTFEDYLHNVVFVPRKTS
CT
FnIII-2
FnIII-3
dmIR_P09208 1051 SKNGSSDKSDGAEGAALDSNAIPNGGATNPSRRRRDVALEPELDDVEGSVLLRHVRSITDDTDAFFEKDDENTYKDEEDL
hsIGF1R_P08069 706 .....ERKRRDVMQVANTT.....MSSRSRNT.........TAADTYN.ITDPEEL
hsA_P06213-2 717 .....RPSKRRSLGDVGNVT.....VAVPTVAA.........F.PNTSS.TSVPTSP
hsB_P06213-1 721 SGTGAEDP.....RPSKRRSLGDVGNVT.....VAVPTVAA.........F.PNTSS.TSVPTSP
FnIII-3
dmIR_P09208 1131 SSNKQTYEVFAKELPPNQTHFVFEKLRHFTRYAIFVVACRERESPSEKLRDTSFKKSLCSDYDTVFQTTKRKKFADIVMDL
hsIGF1R_P08069 742 ETEYPF....FESRVDNKERTVISNLRPFTLYRIDIHSCNHEAEKLGCSASNF.........VFARTMPAEGA...DDI
hsA_P06213-2 754 EEHRPF......EKVVNKESLVISGLRHFTGYRIELQACNQDTPEERCSVAAY.........VSARTMPEAKA...DDI
hsB_P06213-1 766 EEHRPF......EKVVNKESLVISGLRHFTGYRIELQACNQDTPEERCSVAAY.........VSARTMPEAKA...DDI
FnIII-3
dmIR_P09208 1211 KVDLEHANNTESPVRVRWTPPVDPNGEIVTYEVAYKKLQKPDQVEEKKCIPAADFNQTAGY.LIKLNEGLYSFRVRANSIA
hsIGF1R_P08069 805 PGPVTWEPRPENSIFLKWPEPENPNGLILMYEIKYGSQV.E...DQRECVSRQEYRKYGGAKLNRLNPGNYTARIQATSLS
hsA_P06213-2 815 VGPVTHEIFENNVVHLMWQEPKEPNGLIVLYEVSYRRYGDE..ELHLCVSRKHFALERGCRLRGLSPGNYSVRIRATSLA
hsB_P06213-1 827 VGPVTHEIFENNVVHLMWQEPKEPNGLIVLYEVSYRRYGDE..ELHLCVSRKHFALERGCRLRGLSPGNYSVRIRATSLA
FnIII-3
dmIR_P09208 1290 GYGDFTEVEHIKVEPPPSYAKVFFWLLGIG..LAFLIVSLFGYVCYLHKRKVPSNDL.LHMNTEVNPF.....Y.YASMQ
hsIGF1R_P08069 882 GNGSWTDPVFFYVQAKTGYENFIHLIIALPVAVLLIVGGLVIMLYVFHRKRNNSRLGNGVLYASVNPEYFSAADV....V
hsA_P06213-2 893 GNGSWTEPTYFYVTDYLDVPSNIAKIIIGPLIFVFLFSVVIGSIYLFLRKRQP.DGPLGPLYASSNPEYLSASDVFPCSV
hsB_P06213-1 905 GNGSWTEPTYFYVTDYLDVPSNIAKIIIGPLIFVFLFSVVIGSIYLFLRKRQP.DGPLGPLYASSNPEYLSASDVFPCSV
FnIII-3
dmIR_P09208 1359 YIPDDWEVLRENIIQLAPLGQGSFGMVYEGILKSFPPNGVDRECAIKTVNENATDRERTNFLSEASVMKEFDTYHVVRLL
hsIGF1R_P08069 957 YVPDEWEVAREKITMSRELGQGSFGMVYEGVAKGVVKDEPETRVAIKTVNEAASMRERIEFLNEASVMKEFNCHHVVRLL
hsA_P06213-2 972 YVPDEWEVSREKITLLRELGQGSFGMVYEGNARDIIKGEAETRVAVKTVNESASLRERIEFLNEASVMKGFTCHHVVRLL
hsB_P06213-1 984 YVPDEWEVSREKITLLRELGQGSFGMVYEGNARDIIKGEAETRVAVKTVNESASLRERIEFLNEASVMKGFTCHHVVRLL



SI Figure 2. DmIR, hIR-A (hsA) and IR-B (hsB) isoforms, and hIGF-1R (hsIGF1R) sequence alignment. The dmIR-ECD studied here starts span the 291-1309 sequence. Green, yellow and magenta bars mark the span of the dmIR-ECD individual domains.

SI Figure 3. CryoEM data processing workflow and structure solution.



SI Table 1. RMS deviations of the superpositions of the dmIR-ECD and hIR corresponding domains for some representative hIR structures (PDB IDs are given), protomers and the whole ECDs. All superpositions were carried out by the LSQ option in Coot (Emsley *et al*, 2010) on the Ca atoms, except the global superpositions of the whole ECDs for which SSM option in Coot was used. The range of target Cα atoms is given for each type of superposition.

The 827-847 and 881-887 loops in dmIR-ECD were removed prior to the superposition to minimise their potentially misleading bias in the superposition of the 'core' folds of the selected FnIII-1 domains.

Approximate boundaries of the domains in dmIR-ECD and hIR:

	dmIR-ECD domains:	***human IR-A domains:***
L1	288-483	1-154
CR	483-644	155-309
L2	645-804	310-469
FnIII-1	805-927	470-593
FnIII-2	928-1192	594-809
CT	1021-1081	691-720
FnIII-3	1193-1310	810-910

(A) Dynamic protomer FnIII-1 domain superpositions.

Targets: Ala807-Asn925 in dmIR (B-chain) on Glu471 Glu-Asp591 in hIR (mice IR for 7s* PDBs); loops 827-847 and 881-887 deleted for superposition of FnIII-1 and FnIII-1' domains to remove dmIR loops-specific bias. For the IGF-1R the target were Ala807–Asn925 (dmIR B-chain) and Ser461-Ile572 for the IGF-1R (6PYH and 7S0Q (IGF-1R protomer).

6hn5	1.1866 (F chain)	7sl2	1.4645 (A chain)
6sof	1.7768 (A chain)	7sl4	1.2605 (A chain)
6pxv	1.1751 (A chain)	7stj	1.4163 (A chain)
7pg0	1.8630 (B chain)	7stk	1.5234 (A chain)
7pg2	2.9528 (B chain)	7sti	1.6851 (A chain)
7pg3	1.5961 (B chain)	7mqs	1.5243 (E chain)
7pg4	1.5311 (B chain)	6pyh	1.9716 (A chain)

(B) Static protomer FnIII-1 domain superpositions.

Targets: Ala807 - Asn925 in dmIR (A-chain) on Glu471 - Asp591 in hIR (mice IR for 7s* PDBs)

6sof	1.5146 (C chain)	7sl4	1.6013 (B chain)
6hn5	1.3345 (E chain)	7sl7	1.1897 (A chain)
7mqs	1.3040 (F chain)	7sti	1.6024 (B chain)
7pg0	1.6579 (A chain)	7stj	1.5394 (B chain)
7pg2	1.9065 (A chain)	7stk	1.3296 (B chain)
7pg4	1.8980 (A chain)	6pyh	2.3336 (D chain)
7pg3	1.8397 (A chain)	7s0q	1.6135 (A ch
7sl2	1.3535 (B chain)		

(C) Site 1 L1 domains – on the lower arm on dynamic protomer.

Targets: Pro338 – Thr476 in dmIR-ECD (B-chain) and Val7 – Val146 in hIR

7sl2	1.7055 (A chain)	6sof	1.3907 (A chain)
7sl4	1.0463 (A chain)	6hn5	1.0385 (E chain)
7sl7	0.9686 (A chain)	7md4	0.9756 (B chain)
7stk	0.9162 (A chain)		

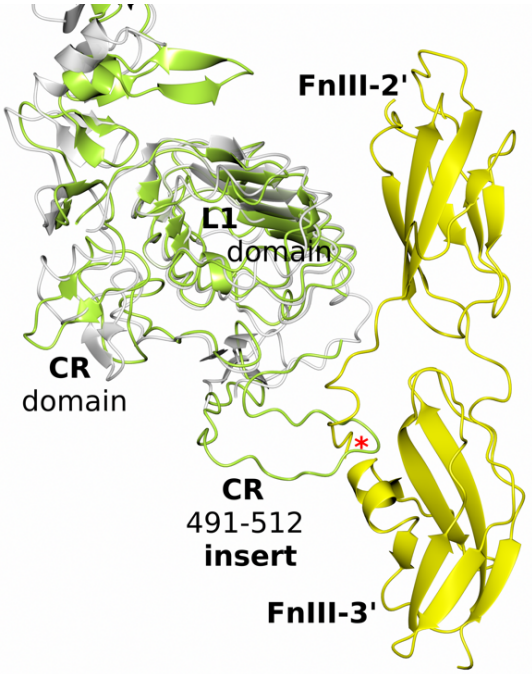
(D) Site 1' L1 domains – on the upper arm on static protomer.

6hn5	1.1042 (E chain)	7sl7	1.1386 (A chain)
6sof	1.4498 (C chain)	7sti	1.1652 (B chain)
7pg0	1.6580 (A chain)	7stj	1.2016 (B chain)
7pg2	1.7206 (A chain)	7stk	1.2498 (B chain)
7pg4	2.0540 (A chain)	7msq	1.4798 (E chain)
7pg3	1.5468 (A chain)	7s0q	1.0295 (A chain)
7sl2	1.1310 (B chain)	6pyh	1.0454 (D chain)
7sl4	1.2421 (B chain)		

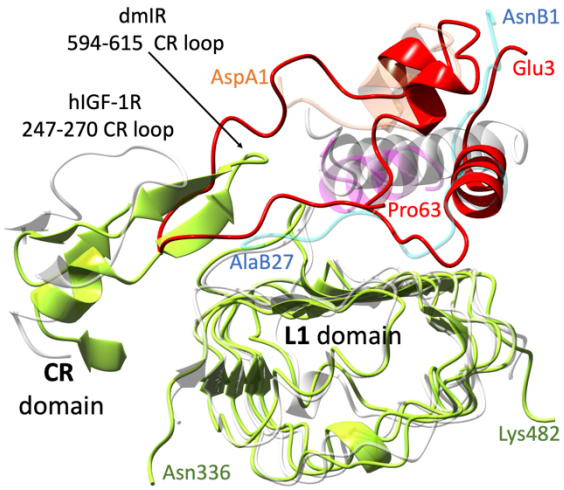
(E) SSM global superpositions of dmIR-ECD and representative hIR (mIR – 7s*) and hIGF-1R (6PYH).

PDB ID, number of atoms in dmIR and in IR (IGF-1R), and number of residues superimposed.

6hn5	4.6359, 1853, 1653, 917	7sl2	4.1117, 1853, 1754, 840
6sof	5.0558, 1853, 1924, 773	7sl4	4.6324, 1853, 1652, 902
6pxv	5.3560, 1853, 1830, 852	7stj	4.5162, 1853, 1672, 886
7pg0	5.2712, 1853, 1742, 1016	7stk	4.3242, 1853, 1726, 860
7pg2	5.1959, 1853, 1700, 1063	7sti	4.3343, 1853, 1678, 855
7pg3	4.3428, 1852, 1721, 884	6pyh	4.6624, 1899, 1648, 705
7pg4	4.6150, 1853, 1677, 913		



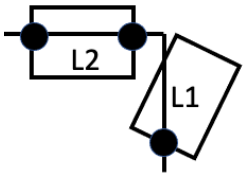
SI Figure 4. Putative steric hinderance caused by dmIR-specific CR 491-512 insert. This putative clash of the dmIR CR 491-512 insert is shown when its corresponding dmIR L1 domain (both in green) is superposed on the L1 domain from the fully-down, insulin-free lower-arm taken here from human IR (6hn5+6hn4) one insulin:IR in site 1 complex (in white). This CR-insert may cause a clash with FnIII-3' domain of the stem of the hIR (here taken from 6hn4 hIR lower part structure, in yellow), and, subsequently with analogous region of dmIR. The red star indicate the overlapping regions.



SI Figure 5. A putative clash of hIGF-1 C-domain with dmIR 594-615 CR loop. dmIR in green, hIGF-1R in grey, hIGF-1 in red, DILP5 A- and B-chains in coral and blue, α–CT in magenta. hIGF-1:hIGF-1R 7S0Q complex structure was selected here due to a better definition of the loop region than in 6PYH.

SI Table 2. Comparison of some angular movements in dmIR-ECD and hIR. Individual domains are depicted by boxes, Cα atoms selected as structural pivots are as black circles.

(A) L2-L1 domains angle:



Cα pivots: 522TrpL1(183Trp) – 660GluL2(Glu329) – 792AspL2(456Asp) (hIR in brackets)

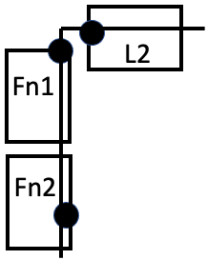
dmIR-ECD:
165.9° down Λ arm (B chain)
108.0° up Γ arm (A chain)

6sof both T arms:
A chain 110.1°
B chain 113.1°

6hn5:

E chain	114.0° up Γ arm
D/F chain	158.8° down Λ arm

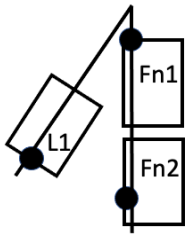
(B) L2-F1-F2 for up Γ arms:



Cα pivots: 696Ala(360Leu) - 908Ile(Thr571) – 938Thr(Val604)

dmIR	A chain	101.9°
6sof	C chain	86.5°
	A chain	90.0°
6hn5	C chain	92.2°
7pg0	A chain	99.0°
7sth	A chain	87.7°
	B chain	87.9°
7stj	B chain	94.0°

(C) L1-F1-F2 for down Λ arms:



Cα pivots: 432LeuL1(104Leu) – 908IleF1(571Thr) – 938ThrF2(604Val); amino acids pivots in the L1 domain were selected for their structural similarity.

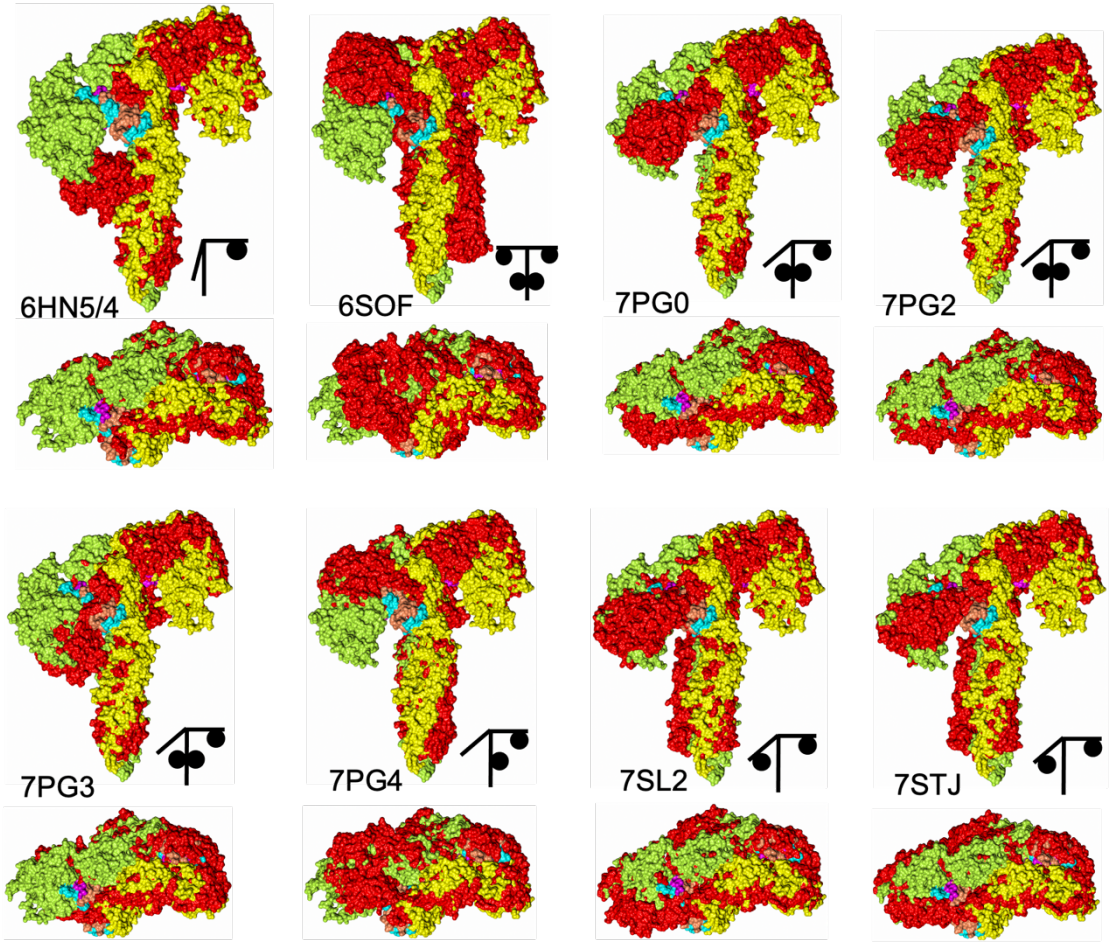
dmIR	A chain	60.4°	
6hn5		22.0°	L1 D chain, F1 F chain, F2 D chain
7pg0	B chain	65.0°	
7stj	A chain	50.9°	
7stk	A chain	53.4°	
7pg3	B chain	35.4°	
7pg4	B chain	72.4°	

Summary of the observed trends.

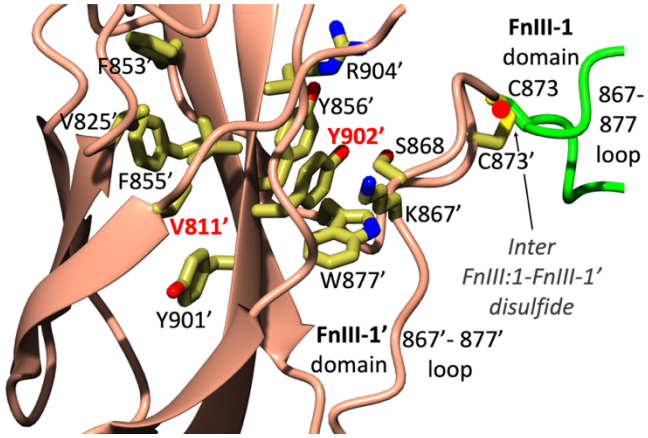
The relative angles between L1-L2 domains, as determined by 522TrpL1(183Trp) – 660GluL2(Glu329) – 792AspL2(456Asp) angle, follow similar trends in all IRs. In dmIR they are 166° for the down Λ arm and 108.0° for the up *Γ* arm. In symmetrical T-form hIR, for example in the 6SOF structure, they are 110°/113° , while in the asymmetric one-insulin hIR complex (PDB ID: 6HN5+6HN4) they are 158.8° in the down Λ arm and 114.0° in the up *Γ* arm. The angles that depicts the 'rectangularity' between the FnIII-domains formed stem of the IRs (FnIII-1-FnIII-2-FnIII-3) and the L2 domain (represented by 695AlaL2(360Leu) – 908IleF1(571Thr) – 938ThrF2(604Val) angle) are within ca. 86° (6SOF) to 99° (7PG0) range in the hIRs, with 102° of the dmIR-ECD in the upper *Γ* arms following this trend.

The largest angular diversity is observed for the L1-FnIII-1-FnIII-2 angle for the down Λ dynamic protomers' arms, which reflects the range of the detachment of their L1 domain from the stem of the IR. As expected, this angle is lowest - 22° - for the fully IR-stem attached L1 domain in one-insulin at the upper site 1 complex (6HN5), attaining subsequently a broad range of values from 35° (7PG3)(3ins/IR), 65° (7PG0)(model 1 – 3 ins/IR), 72° (7PG4)(model 4 with two ins/IR), and 51° and 53° in the hIR structures with site 1 down Λ-like arm insulin. Therefore, the 60° angle of the Λ arm in the DILP5:dmIR-ECD

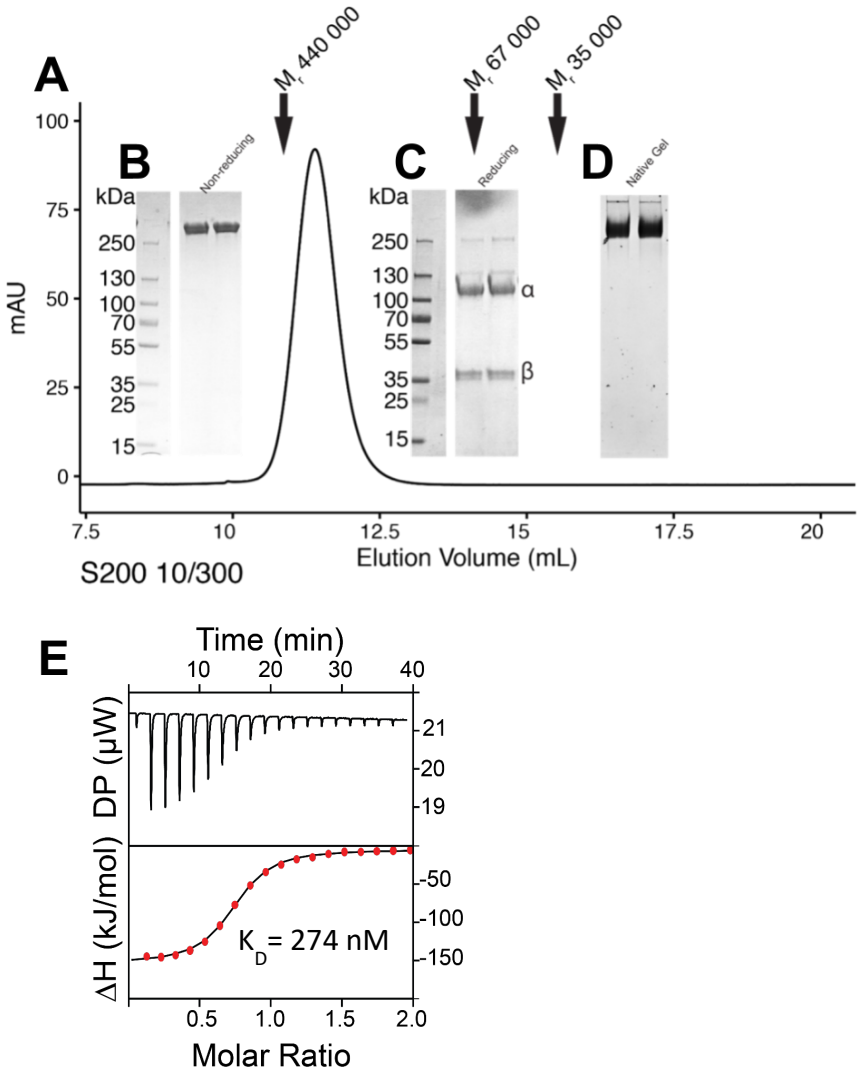
complex falls well within this range. The main – dmIR-specific - 'outlier' here is probably the furthest deviation of the L1-CR-L2 arm from the 'T'-plane of the IR, if measured, for example, between Arg 411(83) from the edge of the L2 as a reference points. Arg411 of the down-arm L2 domain is separated from its human equivalents by ~9-15 Å (after global SSM superpositions of the receptors; otherwise all above geometry indicators are based on specific superpositions targets of the appropriate domains.



SI Figure 6. Global SSM superposition of dmIR-ECD and selected representative hIR (mice IR) structures; colour coding as in **Fig. 5**.



SI Figure 7. The possible impact of Y902C and V811D mutations in dmIR (REF)(see main text). The FnIII-1' and FnIII-1 domains folds are in coral and green, respectively; red dot indicates the C873-C873' inter-protomers disulfide that crosslinks the equivalent FnIII-1 domains. Only the side chains from hydrophobic cavities that surround Y982 and V811 are indicated.



SI Figure 8. Typical purification and size exclusion chromatography of dmIR ecto. (A) Freshly purified dmIR ecto domain analyzed by SEC in 50 mM Tris, 150 mM NaCl, pH 7.4 using a Superdex S200 10/300 GL column (Cytiva). Arrows indicate elution volume of markers 440 000 Da (Ferritin), 67 000 (BSA) and 35 000 β-Lactoglobulin). (B) 4-12% SDS-PAGE of SEC peak containing fractions under non-reducing conditions (C) under reducing conditions and (D) 10% Native gel of SEC peak containing fraction. (E) Typical µITC trace showing titration of DILP-5 hormone into *Drosophila* IR *ecto* domain.