

Supplementary information

Natural linker from patent search

SynLinker Entry		PDB sequence		Linker Information						Structure	Selected Linkers
Accession	Name	Start	End	ID	Residues	Angstrom	Solvent Accessibility	Flexibility	Sequence	Show/Hide	Add/Remove
NP_1	(US4946778A) A helical segment from human hemoglobin Linker 1	-	-	1	22	24.866	-	11.657	AQGTLSPADKTNVKAAWGKVMF	show	<input type="checkbox"/>
NP_10	(US6165476A) Human IgG3 Middle Hinge Linker 2	-	-	1	42	46.633	-	24.576	EPKSCDTPPCPRCEPKSCDTPPCPRCEP KSCDTPPCPRCP	show	<input type="checkbox"/>
NP_11	(US6165476A) Human IgG3M15 Middle Hinge Linker 1	-	-	1	11	17.115	-	5.5152	CDTPPCPRCP	show	<input type="checkbox"/>
NP_12	(US6165476A) Human IgG4 Upper Hinge Linker 1	-	-	1	7	5.544	-	4.9091	ESKYGPP	show	<input type="checkbox"/>
NP_13	(US6165476A) Human IgG4 Middle Hinge Linker 1	-	-	1	5	12.172	-	2.202	CPSCP	show	<input type="checkbox"/>
NP_14	(US6165476A) Human IgG4 Lower Hinge Linker 1	-	-	1	8	17.439	-	4.3232	APEFLGGP	show	<input type="checkbox"/>
NP_15	(US6165476A) Mouse IgG1 Upper Hinge Linker 1	-	-	1	6	15.258	-	2.8737	VPRDCG	show	<input type="checkbox"/>
NP_16	(US6165476A) Mouse IgG1 Middle Hinge Linker 1	-	-	1	7	17.794	-	2.3485	CKPCI CT	show	<input type="checkbox"/>
NP_17	(US6165476A) Mouse IgG1 Lower Hinge Linker 1	-	-	1	6	10.07	-	3.399	VPSEVS	show	<input type="checkbox"/>
NP_18	(US6165476A) Mouse IgG2A Upper Hinge Linker 1	-	-	1	9	20.641	-	5.9091	EPRGPTI KP	show	<input type="checkbox"/>
NP_19	(US6165476A) Mouse IgG2A Middle Hinge Linker 1	-	-	1	7	12.455	-	3.2273	CPPCKCP	show	<input type="checkbox"/>
NP_2	(US6165476A) Human IgG1 Upper Hinge Linker 1	-	-	1	10	21.628	-	6.4192	EPKSCDKTHT	show	<input type="checkbox"/>
NP_20	(US6165476A) Mouse IgG2A Lower Hinge Linker 1	-	-	1	8	20.619	-	4.1919	APNLLGGP	show	<input type="checkbox"/>
NP_21	(US201002217 78A1) Cellulase Linker 1	-	-	1	36	46.054	-	21.126	PGNPTTTVVPPASTSTSRPTSSTSSPVSTP TGQPGG	show	<input type="checkbox"/>
NP_22	(US201002217 78A2) Cellulase Linker 2	-	-	1	35	12.618	-	22.409	PDGGSGNPNPPVSSSTVPSSSTTSSGSSG PTGGT	show	<input type="checkbox"/>

NP_23	(US201002217 78A3) Cellulase Linker 3	-	-	1	24	41.493	-	14.985	GTPNPPASSSTTGSSTPTNPPAG	show	<input type="checkbox"/>
NP_24	(US201002217 78A4) Cellulase Linker 4	-	-	1	33	48.362	-	19.722	PGAGNGGNGGNPPPTTTTSSAPATTTTA SAG	show	<input type="checkbox"/>
NP_25	(US201002217 78A5) Cellulase Linker 5	-	-	1	26	26.232	-	15.576	GGNPPGGNRGTTTTRRPATTGSSPG	show	<input type="checkbox"/>
NP_26	(US201002217 78A6) Cellulase Linker 6	-	-	1	25	43.048	-	13.677	TGTGTGTGTGTGTGTTSSAPAA	show	<input type="checkbox"/>
NP_27	(US201002217 78A7) Cellulase Linker 7	-	-	1	34	44.24	-	22.374	GSSGTPPSNPSSASPTSSTAKPSSTSTAS NPSG	show	<input type="checkbox"/>
NP_28	(US201002217 78A8) Cellulase Linker 8	-	-	1	33	42.871	-	19.157	GTSTGGSTTTTASGTTSTKASTTSTSTST GTG	show	<input type="checkbox"/>
NP_29	(US201002217 78A9) Cellulase Linker 9	-	-	1	33	29.358	-	19.394	TVSSSSVSSSHSSTSTSSSHSSSTPPTQP TGV	show	<input type="checkbox"/>
NP_3	(US6165476A) Human IgG1 Middle Hinge Linker 1	-	-	1	5	9.1	-	2.2172	CPPCP	show	<input type="checkbox"/>
NP_30	(US201002217 78A10) Cellulase Linker 10	-	-	1	48	33.725	-	28.298	PSSGGTSSSSAAPQSTSTKASTTTSVRT TSTATTKTTSSAPAQGTN	show	<input type="checkbox"/>
NP_31	(US201002217 78A11) Cellulase Linker 11	-	-	1	27	30.195	-	17.232	GGNPPGGNPPGTTTTPRPATSTGSSPGP	show	<input type="checkbox"/>
NP_32	(US201002217 78A12) Cellulase Linker 12	-	-	1	41	52.339	-	23.354	ASSSSSTRAASTTSRVSPITTSRSSSATPPP GSTTTRVPPVG	show	<input type="checkbox"/>
NP_33	(US201002217 78A13) Cellulase Linker 13	-	-	1	25	21.021	-	15.177	PPPPASSTTFSTTRRSSTTSSSPS	show	<input type="checkbox"/>
NP_34	(US 20120034160A 1) 5-13 amino acids from the N termini of human Ck and CH1 domains Linker 1	-	-	1	6	14.995	-	3.9444	QPKAAP	show	<input type="checkbox"/>
NP_35	(US 20120034160A 2) 5-13 amino acids from the N termini of human Ck and CH1 domains Linker 2	-	-	1	5	12.76	-	2.1465	TVAAP	show	<input type="checkbox"/>
NP_36	(US 20120034160A 3) 5-13 amino acids from the N termini of human Ck and CH1 domains Linker 3	-	-	1	6	17.231	-	3.9646	ASTKGP	show	<input type="checkbox"/>
NP_37	(US 20120034160A 4) 5-13 amino acids from the N termini of	-	-	1	13	36.388	-	6.9444	QPKAAPSVTLFPP	show	<input type="checkbox"/>

human Ck and CH1 domains Linker 4											
NP_38	(US 20120034160A 5) 5-13 amino acids from the N termini of human Ck and CH1 domains Linker 5	-	-	1	12	32.844	-	4.6919	TVAAPSVFI FPP	show	<input type="checkbox"/>
NP_39	(US 20120034160A 6) 5-13 amino acids from the N termini of human Ck and CH1 domains Linker 6	-	-	1	13	38.002	-	6.899	ASTKGPSVFPLAP	show	<input type="checkbox"/>
NP_4	(US6165476A) Human IgG1 Lower Hinge Linker 1	-	-	1	8	13.381	-	4.4242	APELLGGP	show	<input type="checkbox"/>
NP_5	(US6165476A) Human IgG2 Upper Hinge Linker 1	-	-	1	3	7.083	-	2.4848	ERK	show	<input type="checkbox"/>
NP_6	(US6165476A) Human IgG2 Middle Hinge Linker 1	-	-	1	9	17.959	-	3.3333	CCVECPCP	show	<input type="checkbox"/>
NP_7	(US6165476A) Human IgG2 Lower Hinge Linker 1	-	-	1	7	17.869	-	3.7828	APPVAGP	show	<input type="checkbox"/>
NP_8	(US6165476A) Human IgG3 Upper Hinge Linker 1	-	-	1	12	25.683	-	6.5859	ELKTPLGDTTHT	show	<input type="checkbox"/>
NP_9	(US6165476A) Human IgG3 Middle Hinge Linker 1	-	-	1	5	9.56	-	2.0101	CPRCP	show	<input type="checkbox"/>

Artificial linker from patent search

SynLinker Entry		PDB sequence		Linker Information						Structure	Select Linkers
Accession	Name	Start	End	ID	Residues	Angstrom	Solvent Accessibility	Flexibility	Sequence	Show / Hide	Add / Remove
AP_1	(US4946778 A) Double Linker 1	-	-	1	3	7.018	-	2.0909	PGS	show	<input type="checkbox"/>
AP_10	(US5525491 A) Serine-rich Linker 6	-	-	1	13	11.15	-	9.096	SGSSSSGSSSSGS	show	<input type="checkbox"/>
AP_11	(US5525491 A) Serine-rich Linker 7	-	-	1	19	26.2	-	11.99	SVTVSSSGSSSSGSSSSGS	show	<input type="checkbox"/>
AP_12	(US5856456 A) Linker 1	-	-	1	18	40.278	-	12.707	GSTSGSGKPGSGEGSTKG	show	<input type="checkbox"/>
AP_13	(US5856456 A) Linker 2	-	-	1	18	16.437	-	12.232	GSTSGSGRPGSGEGSTKG	show	<input type="checkbox"/>
AP_14	(WO2012083 424A1) Linker 1	-	-	1	3	6.27	-	1.8737	ATK	show	<input type="checkbox"/>
AP_15	(WO2012083 424A2) Linker 2	-	-	1	3	6.713	-	2.1212	ASK	show	<input type="checkbox"/>
AP_16	(WO2012083 424A3) Linker 3	-	-	1	6	11.199	-	3.9949	ATKASK	show	<input type="checkbox"/>
AP_17	(WO2012083 424A4) Linker 4	-	-	1	7	9.835	-	4.3889	ATKGATK	show	<input type="checkbox"/>
AP_18	(US2013028 0285A1) Modified hinge region of the human CD8 alpha-chain Linker 1	-	-	1	62	33.315	-	30.182	ALSNSI MYFSHFVVPVFLPAKPTTTPAPRPPTPAPT I ASQPLSLRPEASRPAGGAVHTRGLD	show	<input type="checkbox"/>
AP_2	(US4946778 A) Double Linker 2	-	-	1	7	9.931	-	3.7071	I AKAFKN	show	<input type="checkbox"/>
AP_3	(US4946778 A) Single Linker 1	-	-	1	18	31.742	-	10.753	KESGSVSSEQLAQFRSLD	show	<input type="checkbox"/>
AP_4	(US4946778 A) Single Linker 2	-	-	1	14	28.412	-	5.101	VRGSPAI NVAHVHF	show	<input type="checkbox"/>
AP_5	(US5525491 A) Serine-rich Linker 1	-	-	1	5	9.382	-	3.5101	SSSSG	show	<input type="checkbox"/>
AP_6	(US5525491 A) Serine-rich Linker 2	-	-	1	10	16.173	-	7.0202	SSSSGSSSSG	show	<input type="checkbox"/>
AP_7	(US5525491 A) Serine-rich Linker 3	-	-	1	15	18.03	-	10.53	SSSSGSSSSGSSSSG	show	<input type="checkbox"/>
AP_8	(US5525491 A) Serine-rich Linker 4	-	-	1	20	33.484	-	14.04	SSSSGSSSSGSSSSGSSSSG	show	<input type="checkbox"/>
AP_9	(US5525491 A) Serine-rich Linker 5	-	-	1	25	40.363	-	17.551	SSSSGSSSSGSSSSGSSSSGSSSSG	show	<input type="checkbox"/>

TM-align Results

```
*****
*                               TM-align (Version 20170708)                               *
* An algorithm for protein structure alignment and comparison                             *
* Based on statistics:                                                               *
*   0.0 < TM-score < 0.30, random structural similarity                               *
*   0.5 < TM-score < 1.00, in about the same fold                                   *
* Reference: Y Zhang and J Skolnick, Nucl Acids Res 33, 2302-9 (2005)               *
* Please email your comments and suggestions to: zhng@umich.edu                       *
*****

Name of Chain_1: A935253
Name of Chain_2: B935253
Length of Chain_1: 328 residues
Length of Chain_2: 493 residues

Aligned length= 328, RMSD= 0.48, Seq_ID=n_identical/n_aligned= 1.000
TM-score= 0.99481 (if normalized by length of Chain_1)
TM-score= 0.66286 (if normalized by length of Chain_2)
(You should use TM-score normalized by length of the reference protein)
```