Supporting Information

Accurate MALDI-TOF/TOF Sequencing of One-Bead-One-Compound Peptide Libraries, with application to the identification of multi-ligand protein affinity agents using in situ click chemistry screening

Su Seong Lee[*,†], Jaehong Lim[†], Sylvia Tan[†], Junhoe Cha[†], Shi Yun Yeo[†], Heather D. Agnew[‡], and James R. Heath*[*,†,‡]*

[†]Institute of Bioengineering and Nanotechnology, 31 Biopolis Way, The Nanos, Singapore 138669

[‡]Division of Chemistry and Chemical Engineering, Caltech. MC 127-72, Pasadena, CA 911125

sslee@ibn.a-star.edu.sg/heath@caltech.edu

Figure S1. Biotinylation of the biligand for the dot blot.

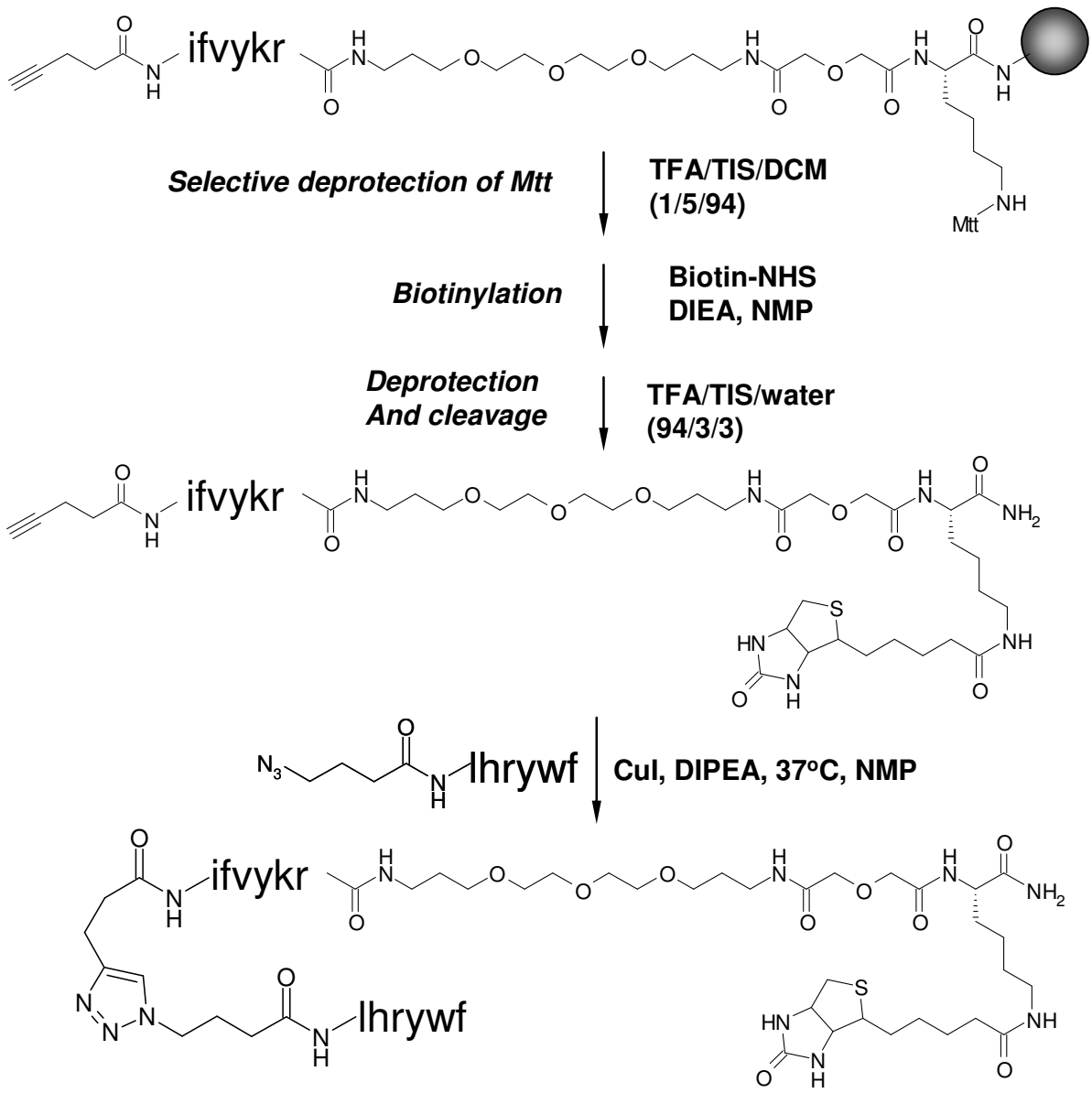


Figure S2. Typical sorting process with COPAS Plus by optimizing gating and sorting region. a) First sorting is accomplished by accepting right side of $\overline{ab}$ to eliminate broken beads and above $\overline{bc}$ to collect only beads associated with the dye labeled protein. The results show that uniformly-sized beads with high red fluorescence are pre-selected for the second sorting. b) The second sorting is accomplished by selecting the region above the line $\overline{de}$ (combination of high red fluorescence and uniform size by running BioSort software (Union Biometrica)) for beads with higher red fluorescence. Typically 50 - 100 beads (0.015 - 0.03 % hit rate) were sorted by the COPAS instrument, starting from 200 mg (*ca.* 300 000 beads).

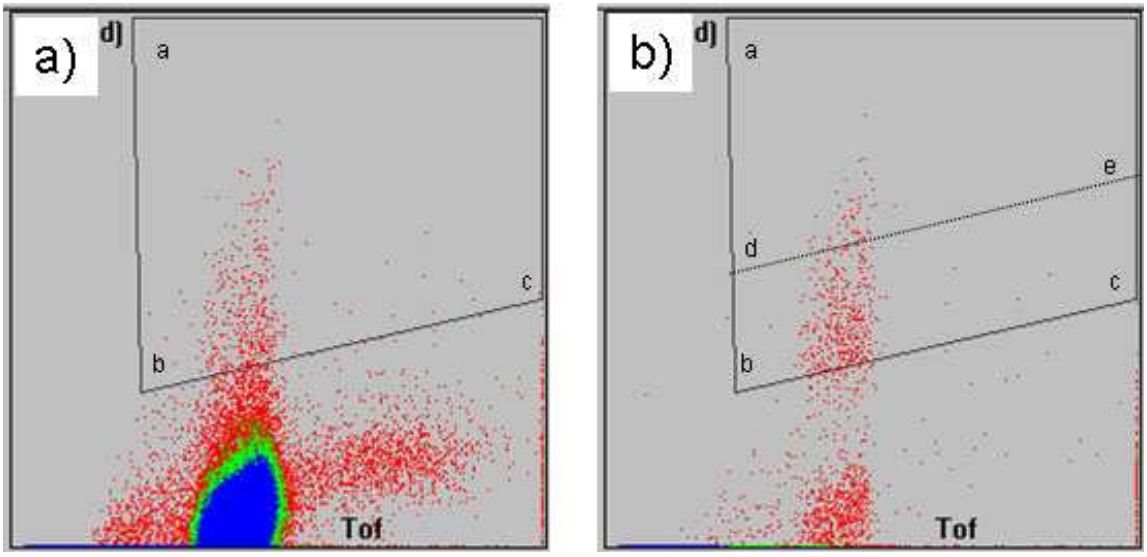


Figure S3. MS spectra of peptides that are not correctly sequenced using standard peak selection algorithms. a) linppm* (m* = homoserine lactone) (entry 41, Table S1) features a prominent peak that arises from a Na^+ adduct to the parent ion. b) gnfetm* (m* = homoserine lactone) (entry 42, Table S1) exhibits a feature that arises from ionization of matrix material, which is very close to the parent mass peak (+6 amu) (* = adventitious peaks from the matrix).

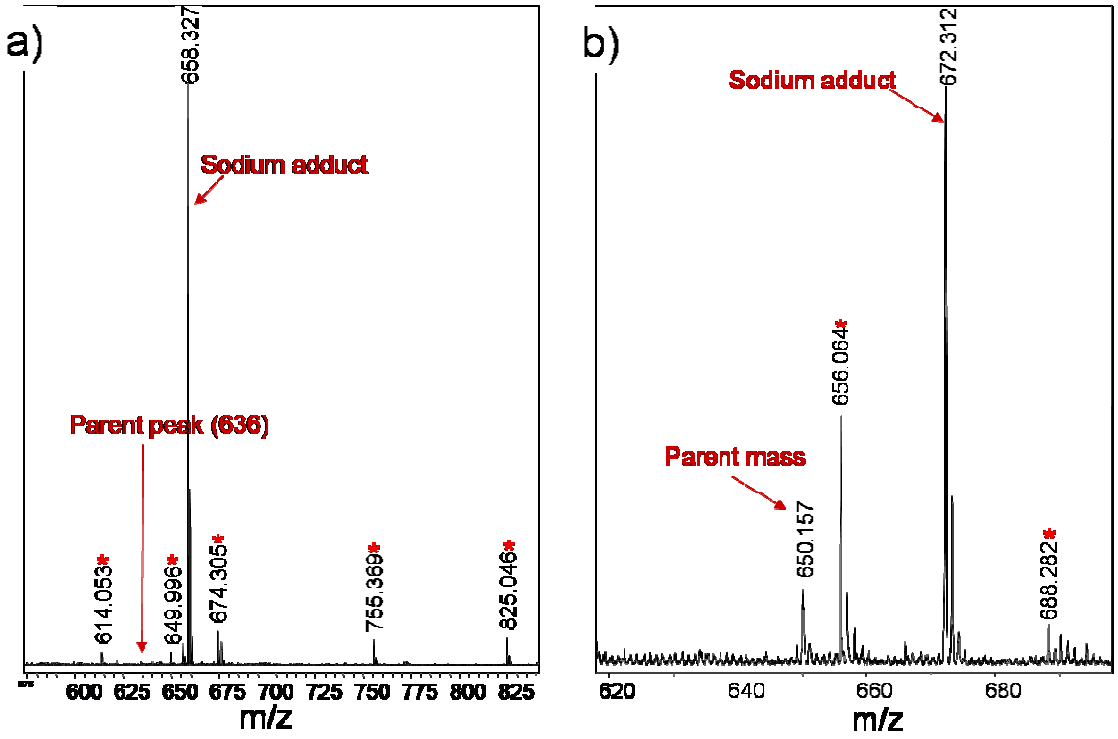


Figure S4. Removal of sodium adducts by washing sample spots on MALDI plate with 10 mM ammonium monobasic phosphate in water (m* = homoserine lactone).

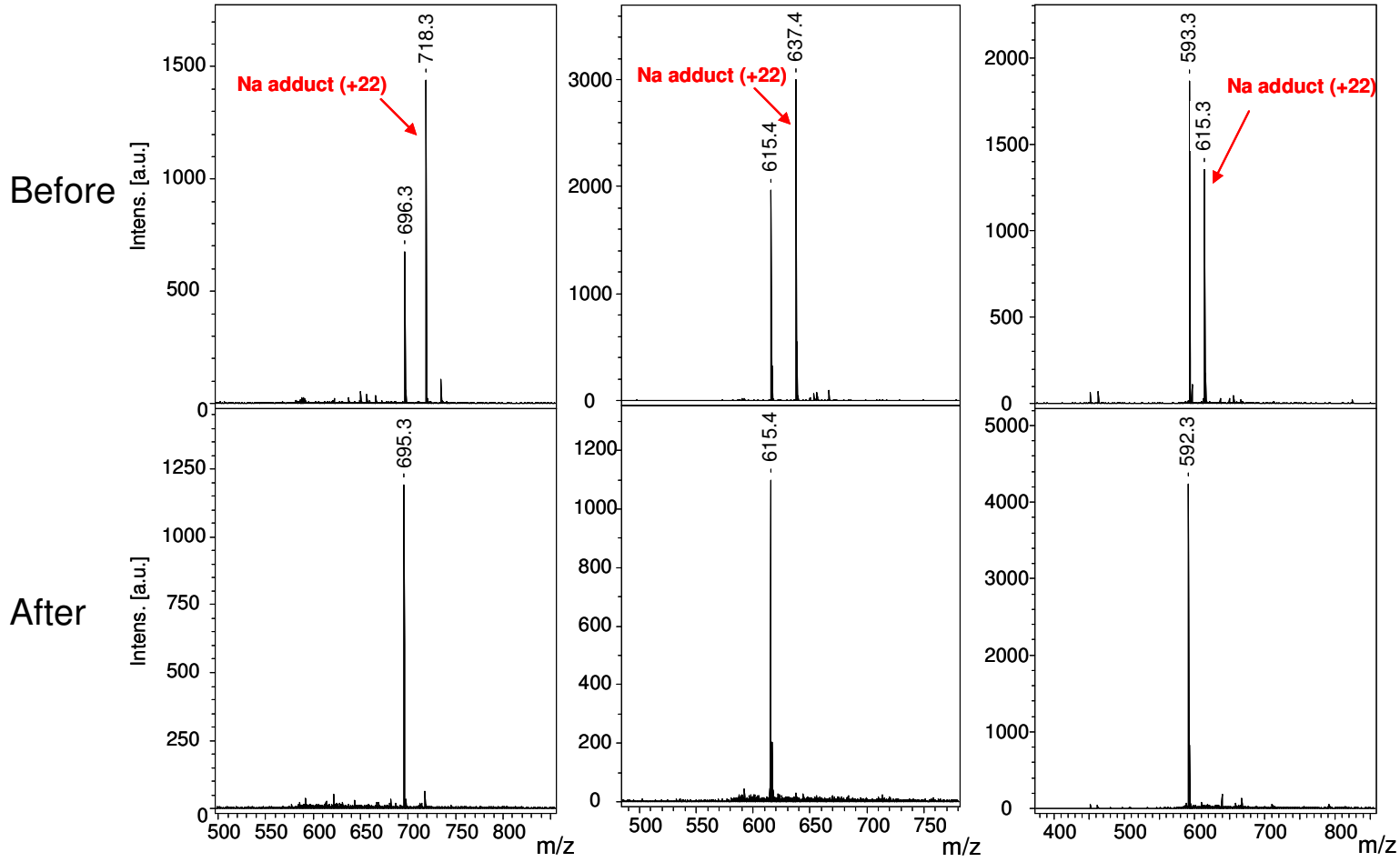


Figure S5. Cyclization of N-terminal amino acids. a) Formation of pyroglutamate from glutamic acid in efrrkm*. b) Formation of pyroglutamate from glutamine in qepfeakm* (m* = homoserine lactone).

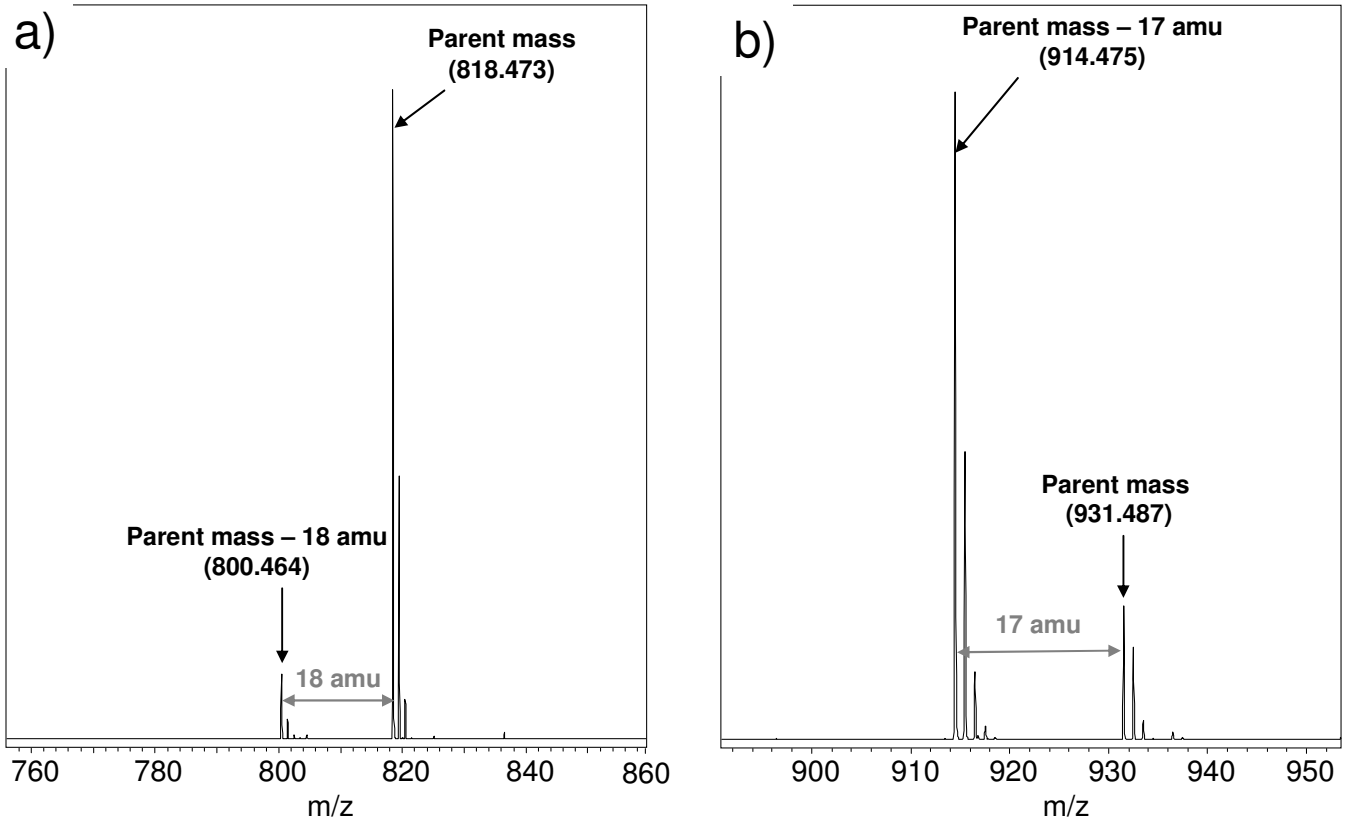


Figure S6. Fragments of rrrrm* (m* = homoserine lactone) in the low mass range (0 – 160). Low mass peaks from R (arginine) are 60, 70, 87, 88, 95, 112, 115, 129, 139, 140 and 157.

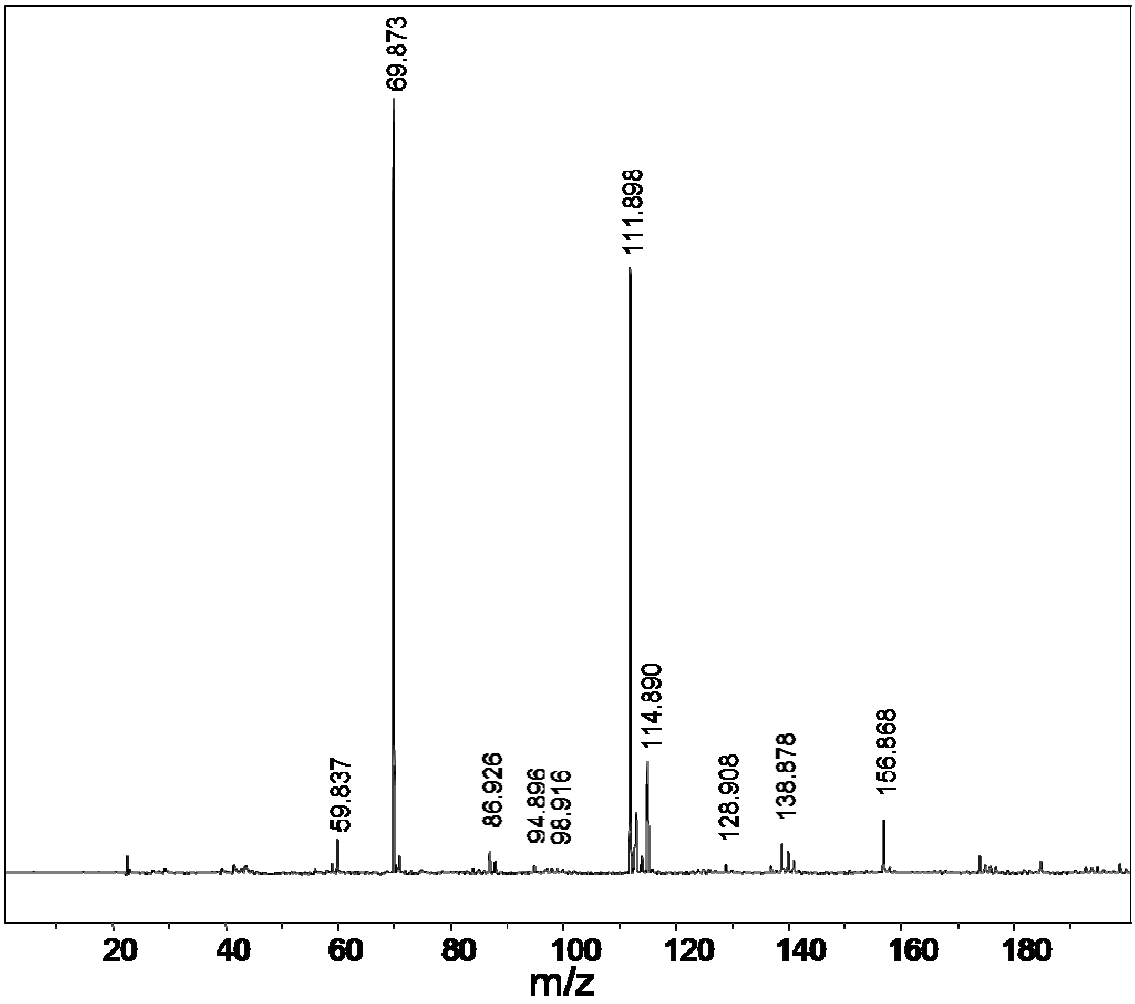


Figure S7. Fragments of rkkkkm* (m* = homoserine lactone) in the low mass range (0 – 160). Low mass peaks from K (lysine) are 84, 101 and 129 (* = the low mass peaks from R).

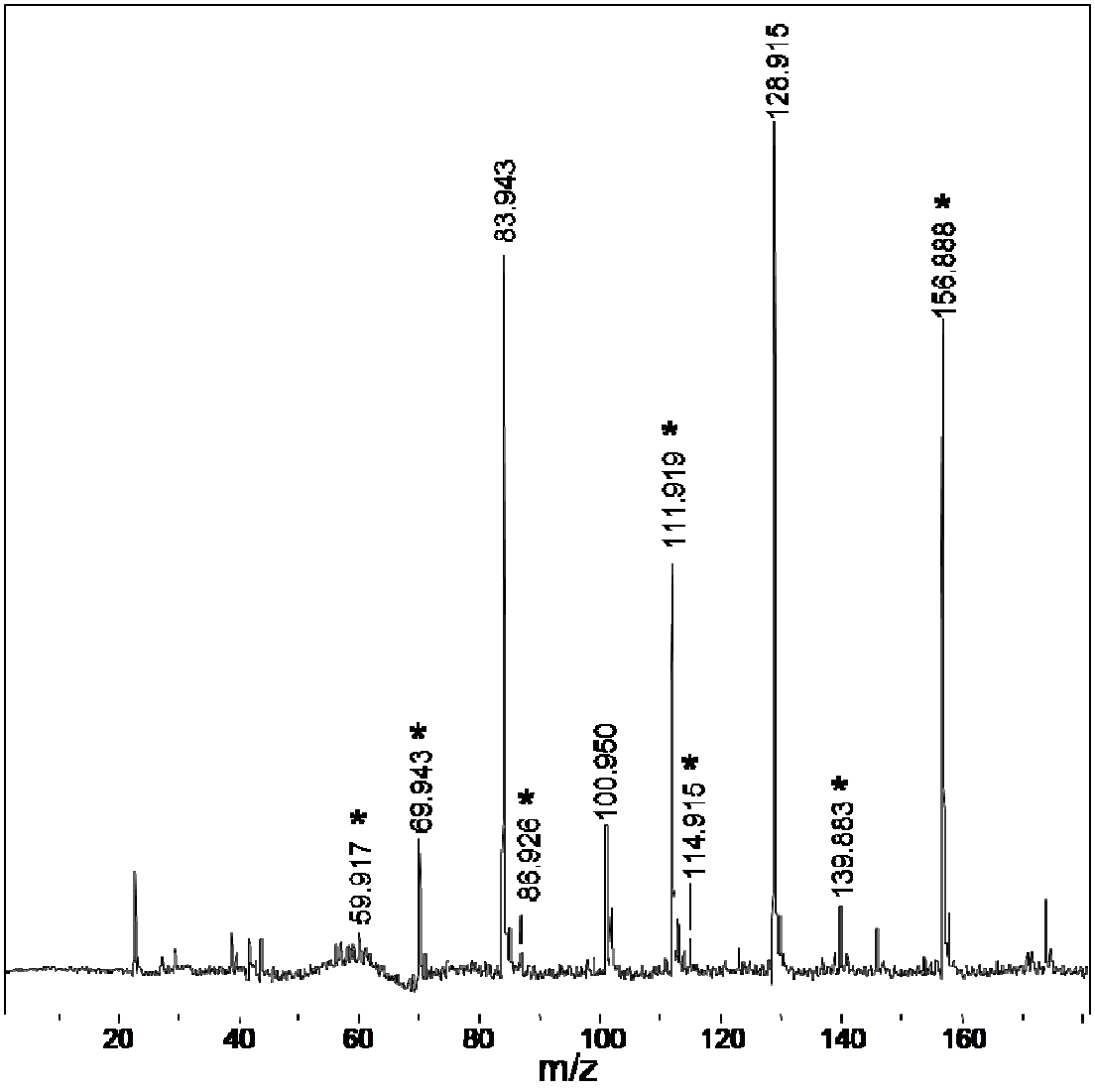


Figure S8. Fragments of wwwwm* (m* = homoserine lactone) in the low mass range (0 – 160). Low mass peaks from W (tryptophan) are 130 and 159.

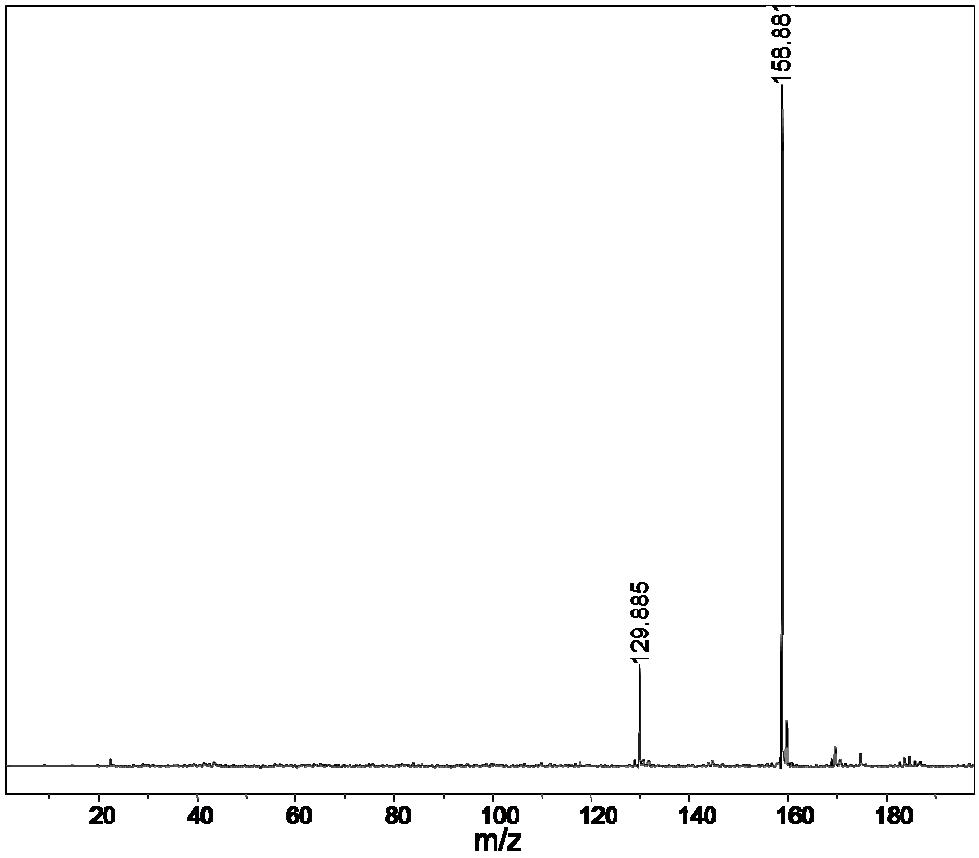


Figure S9. Examples of sequencing peptides that contain isobaric amino acids. Sequencing of a) (l/i)(k/q)h(k/q)rfm* and b) ighkrfm* (m* = homoserine lactone).

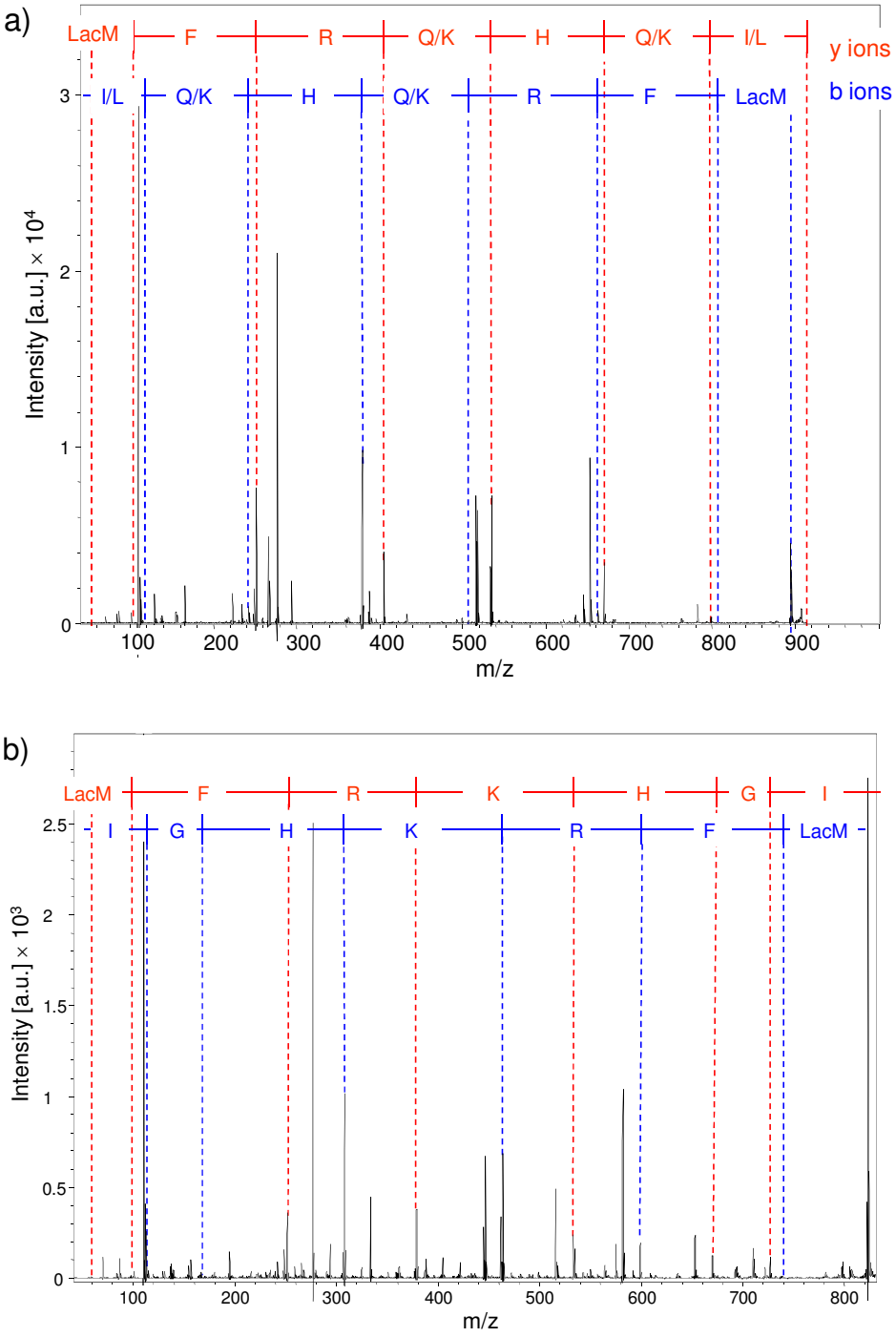
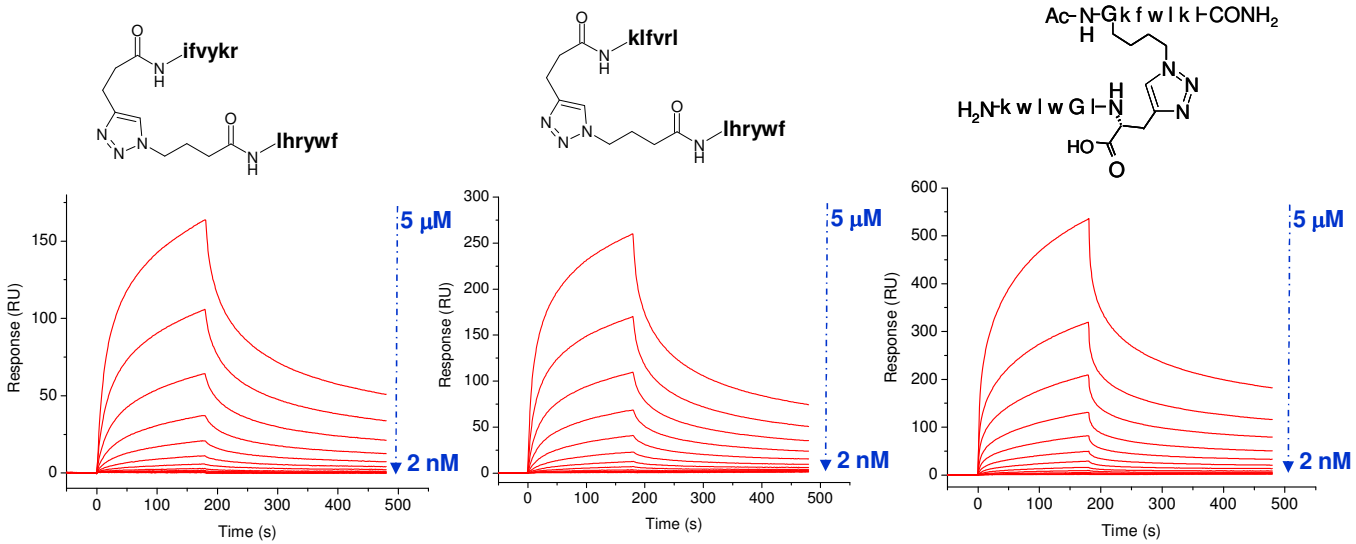


Figure S10. SPR response sensograms for the biligand capture agents (concentration of biligands varied from 2 nM to 5 µM). Kinetic analysis by Biacore T100 Evaluation software (version 2.0.1, Biacore).

Concentration range **(2nM to 5uM)**	ifvykr-lhrywf	klfvrl-lhrywf	Caltech biligand*
Kinetics (2 state model)	1.7	1.3	1.4
Kinetics (Heterogeneous ligand)	K_{D1}= 3.6 K_{D2}= 1.1	K_{D1}= 0.8 K_{D2}= 2.1	K_{D1}= 2.0 K_{D2}= 0.9



* Agnew, H.D.; Rohde, R.D.; Millward S.W.; Nag, A.; Yeo, W.-S.; Hein, J.E.; Pitram S.M.; Tariq, A.A.; Burns, V.M.; Krom, R.J. Fokin, V.V.; Sharpless, K.B.; Heath, J.R. "Iterative in situ Click Chemistry Creates Antibody-like Protein Capture Agents" *Angew. Chem. Int. Ed.* **2009**, 48, 1-5.

Figure S11. Increase of the 70 amu mass peak featured in the presence of r or p in a peptide. (m* = homoserine lactone) (Supporting Information, Table S1)

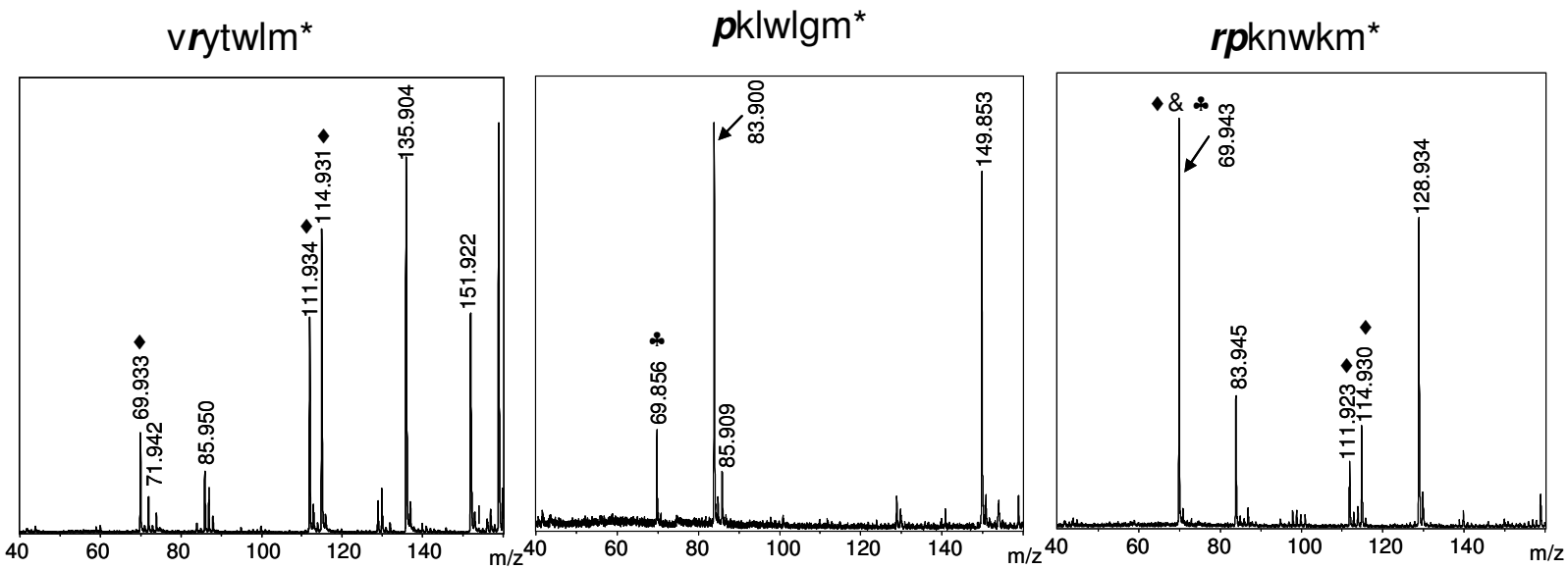


Table S1. Other considerations for improvement of the sequencing algorithm (m* = homoserine lactone).

Manual checking points	Examples
b, y and a ions	
Loss of ammonia from R, K, W, Q, N R shows much clear loss of ammonia	AYRSYM* → **b ions**: AYR – 17, AYRS – 17, AYRSY – 17 and AYRSYM* – 17; **y ions**: RSYM* – 17, YRSYM* – 17 and AYRSYM* – 17
Internal cleavage at P or H	MS/MS spectrum of WLHAGM* contains fragmented ion peaks from HAGM* in addition to those from WLHAGM*.
Relative intensity of low mass peaks	For cases in which p and r existed in the same sequence, the intensity of a 60 amu mass peak noticeably increased by 50 – 150%. (Supporting Information, Figure S11)

Table S2. Sequencing of the halved beads from 48 randomly selected beads (m* = homoserine lactone).

	Observed mass (OM)	Sequence by MS/MS	Sequencing method	Calculated mass (CM)	Sequence by Edman Deg.	Remarks
1	706	hlkllm*	PEAKS	706	hlkii m	
2	713	nptwlm*	PEAKS	713	nptwim	
3	628	lvgrtm*	PEAKS	628	lvgrtm	
4	743	frklpm*	PEAKS	743	frkipm	
5	711	hlvlfm*	PEAKS	711	hlvlfm	
6	742	nwralm*	PEAKS	742	nwraim	
7	696	htlrsm*	PEAKS	696	htirsm	
8	782	fwphlm*	PEAKS	782	fwphim	
9	610	gtvnhm*	PEAKS	610	gtvnhm	
10	752	wfestm*	PEAKS	752	wfestm	
11	671	altfhm*	PEAKS	671	altfhm	
12	914	hrwwfm*	PEAKS	914	hrwwfm	
13	821	wnyqqm*	PEAKS	821	wnyqkm	
14	783	rdhwsm*	PEAKS	783	rdhwsm	
15	784	nwlrlm*	PEAKS	784	nwirlm	
16	733	plqhrm*	PEAKS	733	pikhrm	
17	718	lynllm*	PEAKS	718	lyniim	
18	693	yvnqsm*	PEAKS	693	yvnksm	
19	807	rekeym*	PEAKS	807	rekeym	
20	750	nfnwsm*	PEAKS	750	nfnwsm	
21	799	lykrhm*	PEAKS	799	iykrhm	
22	676	qavylm*	Semi-auto	676	kavyim	
23	644	wldagm*	Semi-auto	644	wldagm	
24	743	drqnqm*	Semi-auto	743	drqnkm	
25	730	hpkhem*	Semi-auto	730	hpkhem	From Edman/MS clues

	Observed mass (OM)	Sequence by MS/MS	Sequencing method	Calculated mass (CM)	Sequence by Edman Deg.	Remarks
26	614	pktegm*	Semi-auto	614	pqtegm	
27	821	lweykm*	Semi-auto	821	lweykm	
28	672	lstrim*	Semi-auto	672	lstrim	
29	873	hyewrm*	Semi-auto	873	hyewrm	From Edamn/MS clues
30	612	tpklam*	Semi-auto	612	tpklam	
31	676	gwadem*	Semi-auto	676	gwadem	
32	640	pddpnm*	Semi-auto	640	pddpnm	
33	720	klfddm*	Semi-auto	720	kddlfm	
34	588	ktaegm*	Semi-auto	588	getkam	
35	685	tnrvlm*	Semi-auto	685	trnvlm	
36	655	lvrgkm*	Semi-auto	655	ivrkgm	
37	675	tlflvm*	Semi-auto	675	vdfivm	
38	619	hsfhvm*	PEAKS	597	hsltgm	OM – CM = 22
39	665	nakysm*	PEAKS	643	naqvem	OM – CM = 22
40	616			594	pikpgm	OM – CM = 22
41	658	srpesm*	PEAKS	636	linppm	OM – CM = 22
42	672			650	gnfetm	OM – CM = 22
43	694	ktphem*	PEAKS	672	tqppfm	OM – CM = 22
44	750	vllhwm*	Semi-auto	712	vilvwm	
45	No peak			689	gnewtm	
46	No peak			804	ewsnwm	
47	No peak			591	glkdgm	
48	No peak			555	vgsipm	

Table S3. Sequencing of the halved beads from 28 randomly selected beads by MALDI-TOF/TOF and Edman degradation. No differentiation of isobaric amino acids for MALDI-TOF/TOF, *i.e.*, k = q, l = i (m* = homoserine lactone).

Sample	m/z	MALDI-TOF/TOF	Edman degradation
1	669.366	dvdhtm*	dvdhtm
2	696.316	Ysntem*	ysntem
3	No observation of parent mass	-	not clean
4	679.4	lyetam*	iyetam
5	628.394	**sk**tplm*	**ks**tpim
6	666.361	wlhagm*	wlhagm
7	719.439	evlylm*	evlyim
8	766.425	arnhwm*	arnhwm
9	805.444	lytyym*	iytyym
10	755.462	wtprlm*	wtprlm
11	682.357	hegqem*	hegqem
12	802.444	wswlkm*	wswikm
13	723.427	ftlpym*	ftlpym
14	670.44	**el**ravm*	**rle**avm
15	651.405	hnklgm*	hnkigm
16	615.42	akksvm*	akksvm
17	593.285	aphadm*	aphadm
18	787.318	**efy**vfm*	**fye**vfm
19	647.405	fnkvgm*	fnkvgm
20	718.375	hpwvpm*	hpwvpm
21	857.448	lw**nw**rm*	iw**wn**rm
22	742.373	ayrsym*	ayrsym
23	800.366	nnkwrm*	nnkwrm
24	718.363	wlgpym*	wlgpym
25	747.344	eswrsm*	eswrsm
26	804.399	hrfvym*	hrfvym
27	757.475	wrtvlm*	wrtvim
28	689.385	afrrgm*	afrrgm

Table S4. Sequences of the hit beads from the screening of a focused hexamer library against bCAII (homoserine lactone at C-terminus).

m/z	Sequence					
972.551	k	i	y	r	f	y
1004.63	k	i	r	y	w	r
911.531	i	q	h	k	r	f
953.546	k	y	f	k	v	w
1045.558	y	k	r	y	w	f
1071.568	w	r	w	r	v	w
958.547	v	k	y	r	f	y
999.57	r	r	f	q	f	y
990.613	l	h	r	y	r	y
1036.54	w	r	w	r	y	s
1077.582	w	r	w	y	k	r
1032.6	l	r	w	r	y	r
983.605	l	h	r	y	r	r
989.57	y	r	w	l	r	l
1045.555	y	r	w	k	f	y
936.558	y	l	v	f	r	r
958.67	k	i	f	r	r	r
1049.634	r	r	w	r	r	h
1002.645	r	l	r	r	y	r
1013.546	l	h	r	y	w	r
1019.486	r	q	h	w	y	f
1051.677	k	r	w	h	w	r
1019.532	w	r	r	y	k	k
1000.53	r	q	w	h	y	k
867.584	k	v	l	r	r	l
1061.596	w	r	w	r	f	k
1014.625	w	r	w	k	k	k
875.559	r	v	y	v	q	k
955.619	l	h	r	y	r	k
1004.536	l	h	r	y	w	f
1013.523	l	h	r	y	w	r
1000.537	r	k	w	h	y	q
946.602	v	w	r	f	k	k
952.572	r	y	f	k	q	k
1022.599	w	r	w	h	v	r
955.559	l	h	r	y	r	q
1075.627	w	r	r	y	r	r
1008.604	r	q	r	r	f	y

m/z	Sequence					
987.567	l	r	w	h	r	h
1000.569	l	r	y	r	f	y
1039.554	y	l	r	y	w	r
1004.494	l	h	r	y	w	f
940.547	l	h	r	y	l	r
1020.544	l	h	r	y	w	y
1013.672	l	h	r	y	w	r
962.526	w	k	q	v	y	r
1093.538	w	r	y	h	y	w
1135.553	w	r	y	w	r	w
1013.56	l	h	r	y	w	r
1000.474	w	r	h	q	y	k
985.51	l	h	r	y	k	w
1049.553	w	r	w	k	y	k
985.53	l	h	r	y	w	k
1005.538	y	h	r	y	k	r
985.596	w	r	w	k	k	v
938.565	l	y	i	r	y	k
985.56	l	h	r	y	w	k
990.528	y	r	y	r	i	h
1004.502	l	h	r	y	w	f
1012.63	r	w	i	w	r	l

Table S5. Sequences of hit beads from *in situ* screening of a comprehensive hexamer library (200 mg x 3, P* = 5-pentynoic acid, m* = homoserine lactone).

1	P*	t	f	l	v	r	k	m*		1	P*	v	v	a	l	n	k	m*		1	P*	a	a	f	k	r	l	m*	
2	P*	n	f	f	l	r	k	m*		2	P*	p	r	r	w	k	f	m*		2	P*	k	r	v	w	v	k	m*	
3	P*	h	r	v	w	v	w	m*		3	P*	t	f	y	a	r	s	m*		3	P*	f	f	r	y	t	r	m*	
4	P*	t	l	l	l	f	k	m*		4	P*	l	f	t	r	l	a	m*		4	P*	i	y	f	y	k	f	m*	
5	P*	f	l	k	v	k	s	m*		5	P*	w	k	q	w	k	r	m*		5	P*	t	r	l	f	v	k	m*	
6	P*	p	l	l	t	k	h	m*		6	P*	l	f	y	f	k	p	m*		6	P*	i	f	i	k	r	k	m*	
7	P*	i	i	r	k	f	y	m*		7	P*	i	i	v	f	s	r	m*		7	P*	r	n	t	v	l	w	m*	
8	P*	f	f	y	f	k	k	m*		8	P*	r	i	k	w	w	a	m*		8	P*	h	l	v	v	f	r	m*	
9	P*	n	v	k	f	w	r	m*		9	P*	v	f	y	t	s	k	m*		9	P*	l	f	y	k	f	n	m*	
10	P*	v	k	f	f	r	l	m*		10	P*	i	r	i	y	i	r	m*		10	P*	e	k	v	w	f	r	m*	
11	P*	v	f	y	r	l	s	m*		11	P*	a	i	i	r	r	p	m*		11	P*	v	f	l	r	f	n	m*	
12	P*	n	n	k	f	v	r	m*		12	P*	n	f	y	a	r	v	m*		12	P*	i	s	i	f	f	r	m*	
13	P*	a	v	k	y	l	y	m*		13	P*	f	f	y	r	n	s	m*		13	P*	l	r	v	w	v	w	m*	
14	P*	r	w	f	v	f	k	m*		14	P*	k	h	w	r	y	r	m*		14	P*	h	r	v	w	v	g	m*	
15	P*	i	v	w	f	n	r	m*		15	P*	l	v	l	i	l	r	m*		15	P*	k	a	l	w	l	k	m*	
16	P*	f	f	f	k	s	h	m*		17	P*	v	f	f	q	r	w	m*		16	P*	k	k	f	v	r	v	m*	
17	P*	f	l	k	f	v	k	m*		18	P*	i	f	f	s	r	q	m*		17	P*	v	r	f	y	l	r	m*	
18	P*	v	a	f	n	k	f	m*		20	P*	i	a	f	f	q	f	m*		18	P*	a	r	i	f	f	t	m*	
19	P*	p	l	f	f	p	h	m*		21	P*	k	l	l	f	h	i	m*		19	P*	r	l	t	y	v	n	m*	
20	P*	y	f	y	v	k	k	m*		23	P*	v	n	f	v	k	l	m*		20	P*	k	k	l	f	f	k	m*	
21	P*	r	f	r	y	y	k	m*		24	P*	f	k	l	a	n	r	m*		21	P*	r	f	y	n	r	d	m*	
22	P*	a	v	l	f	n	r	m*		25	P*	i	f	r	y	t	r	m*		22	P*	r	f	y	f	y	k	m*	
23	P*	v	a	f	f	k	v	m*		26	P*	n	l	r	i	r	a	m*		23	P*	f	l	r	v	n	w	m*	
24	P*	v	l	v	l	k	k	m*		27	P*	y	r	i	f	a	t	m*		24	P*	y	l	v	y	r	w	m*	
25	P*	n	a	l	w	f	r	m*		28	P*	t	t	f	i	w	k	m*		25	P*	l	f	k	r	v	l	m*	
26	P*	w	t	n	h	p	s	m*		30	P*	v	i	k	i	t	r	m*		26	P*	t	s	y	f	h	p	m*	
27	P*	k	h	w	k	w	r	m*		31	P*	f	v	f	l	k	k	m*		27	P*	v	l	l	i	v	r	m*	
28	P*	v	l	f	y	r	f	m*		33	P*	v	v	l	t	n	k	m*		28	P*	e	r	f	w	l	r	m*	
29	P*	w	n	y	f	v	r	m*		34	P*	t	r	v	w	a	w	m*		29	P*	k	r	r	f	w	k	m*	
30	P*	k	y	f	l	g	f	m*		35	P*	i	l	l	n	k	y	m*		30	P*	r	f	y	f	r	v	m*	
31	P*	y	k	i	f	v	r	m*		36	P*	t	r	v	w	l	w	m*		31	P*	f	l	v	w	r	y	m*	
32	P*	k	l	v	f	s	r	m*		37	P*	a	k	y	i	r	i	m*		32	P*	r	t	g	s	f	l	m*	
33	P*	l	f	f	f	v	r	m*		38	P*	d	r	t	t	d	w	m*		33	P*	v	v	a	i	n	r	m*	
34	P*	l	s	l	f	f	r	m*		39	P*	g	k	y	k	w	w	m*		34	P*	g	k	l	f	l	k	m*	
35	P*	a	k	f	w	l	g	m*		40	P*	i	f	y	f	q	r	m*		35	P*	y	f	r	v	f	k	m*	
36	P*	i	i	r	y	f	i	m*		41	P*	w	f	y	k	k	p	m*		36	P*	f	g	f	w	y	r	m*	
37	P*	p	p	w	h	p	d	m*		42	P*	k	l	k	w	r	f	m*		37	P*	l	f	r	f	r	k	m*	
38	P*	v	l	f	v	r	l	m*		43	P*	r	n	t	a	l	y	m*		38	P*	r	n	l	k	k	y	m*	
39	P*	h	l	v	f	k	k	m*		44	P*	f	i	r	y	f	i	m*		39	P*	l	r	r	y	k	r	m*	
40	P*	w	f	r	r	k	l	m*		45	P*	h	r	l	f	l	k	m*		40	P*	k	l	w	f	t	r	m*	
41	P*	l	s	v	w	l	r	m*		46	P*	l	r	k	v	l	w	m*		41	P*	y	k	f	k	y	r	m*	
42	P*	v	k	l	f	f	r	m*		47	P*	w	r	t	w	w	k	m*		42	P*	p	r	f	f	k	k	m*	
43	P*	v	l	k	l	v	k	m*		48	P*	w	w	l	k	f	k	m*		43	P*	k	r	r	k	y	a	m*	
44	P*	k	f	y	l	r	r	m*		49	P*	v	n	t	l	k	l	m*											
45	P*	l	r	v	w	l	r	m*		50	P*	i	f	v	s	k	r	m*											
46	P*	p	v	a	w	v	r	m*		51	P*	l	f	y	f	k	t	m*											
47	P*	k	y	f	v	k	f	m*		52	P*	i	r	y	r	w	k	m*											
48	P*	r	y	k	k	r	s	m*		53	P*	t	f	i	v	r	k	m*											
										54	P*	d	r	w	k	r	k	m*											
										55	P*	r	r	a	r	r	y	m*											

Table S6. Sequences of hit beads form *in situ* screening of a focused hexamer library (200 mg x 2, P* = 5-pentynoic acid, m* = homoserine lactone).

1	P*	i	f	v	y	r	k	m*		1	P*	v	f	l	y	k	r	m*
2	P*	v	f	v	y	k	q	m*		2	P*	i	f	i	y	r	k	m*
3	P*	v	f	v	y	k	r	m*		3	P*	k	i	f	v	r	y	m*
4	P*	i	f	i	y	k	k	m*		4	P*	l	k	l	w	f	k	m*
5	P*	i	f	i	y	k	r	m*		5	P*	i	k	f	v	r	y	m*
6	P*	f	k	l	w	v	k	m*		6	P*	f	f	f	y	k	k	m*
7	P*	v	f	f	y	k	k	m*		7	P*	k	l	f	v	r	l	m*
8	P*	l	f	f	y	r	k	m*		8	P*	i	k	i	w	l	k	m*
9	P*	l	r	l	w	f	k	m*		9	P*	v	k	v	f	l	r	m*
10	P*	l	r	l	w	f	k	m*		10	P*	v	k	v	f	f	r	m*
11	P*	l	v	f	y	k	r	m*		11	P*	f	i	k	f	v	r	m*
12	P*	l	k	f	v	r	y	m*		12	P*	k	v	f	v	r	y	m*
13	P*	f	f	r	y	v	k	m*		13	P*	f	f	f	y	k	r	m*
14	P*	i	r	i	f	v	r	m*		14	P*	v	v	f	v	r	r	m*
15	P*	k	l	f	v	r	l	m*		15	P*	v	k	f	w	f	k	m*
16	P*	i	f	i	y	k	r	m*		16	P*	f	r	l	w	v	k	m*
17	P*	v	k	i	f	v	r	m*		17	P*	i	r	l	w	f	k	m*
18	P*	i	f	f	y	k	k	m*		18	P*	r	f	f	y	k	r	m*
19	P*	l	r	l	w	l	k	m*		19	P*	l	k	i	w	l	r	m*
20	P*	i	k	f	v	r	y	m*		20	P*	v	i	f	l	r	s	m*
21	P*	v	n	l	f	l	r	m*		21	P*	v	k	v	f	l	r	m*
22	P*	l	r	l	w	l	k	m*		22	P*	k	f	l	r	f	y	m*
23	P*	i	r	i	f	f	k	m*		23	P*	v	k	v	f	l	r	m*
24	P*	l	r	l	w	f	k	m*		24	P*	i	k	f	v	r	y	m*
25	P*	l	k	l	w	l	k	m*		25	P*	k	f	i	y	k	r	m*
26	P*	v	a	f	f	r	i	m*		26	P*	f	f	k	y	v	k	m*
27	P*	f	f	k	y	v	k	m*		27	P*	i	k	v	f	v	r	m*
28	P*	l	k	l	w	l	k	m*		28	P*	i	k	v	f	v	r	m*
29	P*	l	k	v	y	l	r	m*										
30	P*	k	i	f	v	r	i	m*										