Supporting Information

Over-Represented DNA Libraries Containing Two Hydrophobic (Het)aryl-Linked Nucleotides for Expedient Single-Round Selection of a Specific Aptamer Targeting Human Insulin Receptor

Pablo Alberto Franco-Urquijo, Marek Ondruš, Jaroslav Kurfürst, Jana Škerlová, Irena Selicharová, Lucie Mužíková Čechová, Hana Šváchová, Alena Semerádtová, Anatolij Filimoněnko, Adéla Fejfarová, Jiří Homola, Tomáš Kouba and Michal Hocek

Table of contents

1. General remarks

Sequences of natural templates, primers and candidates were purchased from Generi Biotech (Czech Republic). Randomized DNA pools were obtained from Integrated DNA Technologies (IDT, USA). Natural nucleoside triphosphates (dATP, dTTP, dCTP, dGTP), Dynabeads MyOne Streptavidin C1, Dynabeads His-Tag Isolation & Pulldown, Human Serum Albumin (HSA), Prothrombin, 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC), Sulfo-NHS (N-hydroxysulfosuccinimide) and Dextrane Sulfate sodium salt from Leuconostoc spp. were purchased from ThermoFisher Scientific. KOD XL DNA polymerase was purchased from Merck (Sigma Aldrich). Primer Extension (PEX) and Polymerase chain reaction (PCR) reactions were performed in VWR thermal cycler (Doppio), Quantitative Polymerase chain reactions (qPCR) with CFX96 Real-time System thermal cycler (Bio-Rad). When required, PEX or PCR products were purified using the Monarch PCR & DNA Purification Kit (New England Biolabs, NEB) and QIAquick PCR Purification Kit (Qiagen). Purified product concentrations were measured by UV-Vis spectra at room temperature (r.t.) on Nanophotometer® N60 (IMPLEN) and concentrated on CentriVap Vacuum Concentrator system (Labconco). Samples from PEX reactions were separated by 12.5% polyacrylamide gel electrophoresis (PAGE, acrylamide/bisacrylamide 19:1,25% urea [v/v]) with 1X TBE buffer (420 mA, 1 h) using stop solution (95% [v/v] formamide, 0.5 mM EDTA, 0.025% [w/v] bromophenol blue, 0.025% [w/v] xylene cyanol FF, 0.025% [w/v] SDS in Milli-Q water). PEX, PCR reactions, λ exonuclease digestions, and single-strand separations were separated by native 3% agarose gel (Serva) electrophoresis with 0.5X TBE buffer (120 V, 70 min) using 6X DNA Gel Loading Dye (ThermoFisher Scientific). All gels were analyzed by GelRed or fluorescence imaging using Typhoon FLA 9500 (GE Healthcare Life Sciences) by comparing the migration to a custom-made FAM-/Cy5-labeled DNA ladder or Low Range DNA Ladder (Invitrogen). His-Tagged Human Insulin Receptor (HIR, His28-Lys944, Cat. No. INR-H52Ha), His-Tagged Human IGF-1R (IGF-1R, Glu31-Asn932, Cat. No. IGR-H5229), His-Tagged Cat Insulin Receptor (CIR, His28-Ile956, Cat. No. INR-C52H3) and Dog Insulin Receptor (DIR, His28-Ile956, Cat. No. INR-D52H3) were acquired from ACROBiosystems. His-tagged peptide (HHHHHH) was custom-made in-house. Dulbecco´s PBS buffer containing 1.5 mM KH_2PO_4, 8.1 mM Na_2HPO_4, 2.7 mM KCl, 137 mM NaCl was purchased from BioConcept.

1.1 Oligonucleotides used in this study

Table S1. Sequences of natural templates, primers, libraries and candidates used in this study.

ON title	Sequence (5'→3')	nt
L1_temp[b]	AGTCACACAGAAGTGACGTC-N15-ACGAAGGCTGTATAGATGCG	55
L2_temp[b]	GACATCATGAGAGACATCGC-N15-GAGATACAGTGAGCGTCATG	55
L3_temp[b,d]	GTAGACACATCTGACGAGCAGACATCATGAGAGACATCGCTGTTACGTTAAACGCGAGATACAGTGAGCGTCATGGAGACGCTGATGAGATTAGC	95
Rev1[a]	CGCATCTATACAGCCTTCGT	20
Rev2[a]	CATGACGCTCACTGTATCTC	20
Rev3[a,c]	GCTAATCTCATCAGCGTCTC	20
Fw3[b]	GTAGACACATCTGACGAGCA	20
Rev1_adapter	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCGCATCTATACAGCCTTCGT	53
Fw1_adapter	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGTCACACAGAAGTGACGTC	54
Rev2_adapter	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCATGACGCTCACTGTATCTC	53
Fw2_adapter	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACATCATGAGAGACATCGC	54
Rev3_adapter	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTAATCTCATCAGCGTCTC	53
Fw3_adapter	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTAGACACATCTGACGAGCA	54
L1_c1	CGCATCTATACAGCCTTCGTATGACTACGTACTAGGACGTCACTTCTGTGTGACT	55
L1_c2	CGCATCTATACAGCCTTCGTGTCATACCATCACCGGACGTCACTTCTGTGTGACT	55
L1_c3	CGCATCTATACAGCCTTCGTCTGCAAGCGGATACAGACGTCACTTCTGTGTGACT	55
L1_c4	CGCATCTATACAGCCTTCGTAGGGGGGGTCAGAGAGACGTCACTTCTGTGTGACT	55
L1_c5	CGCATCTATACAGCCTTCGTCAACTTCCGTCCTATGACGTCACTTCTGTGTGACT	55
L2_c1_temp[b]	AGTCACACAGAAGTGACGTCTTCGAGATAGGTCGAACGAAGGCTGTATAGATGCG	55
L2_c2_temp[b]	AGTCACACAGAAGTGACGTCCCATTAGATGTAAGGACGAAGGCTGTATAGATGCG	55
L2_c3_temp[b]	AGTCACACAGAAGTGACGTCTGCAAACGAAGTCTGACGAAGGCTGTATAGATGCG	55
L2_c4_temp[b]	AGTCACACAGAAGTGACGTCGCCCGTAGATGTCTAACGAAGGCTGTATAGATGCG	55
L2_c5_temp[b]	AGTCACACAGAAGTGACGTCTTACGTTAAAGCGGCACGAAGGCTGTATAGATGCG	55
L3_c1_temp[b]	AGTCACACAGAAGTGACGTCGCGAAGTGTAGATGCACGAAGGCTGTATAGATGCG	55

L3_c2_temp[b]	AGTCACACAGAAGTGACGTCTCGAATATCGGGGGCACGAAGGCT GTATAGATGCG	55
L3_c3_temp[b]	AGTCACACAGAAGTGACGTCCCTAGAATGGACCCTACGAAGGCTG TATAGATGCG	55
L3_c4_temp[b]	AGTCACACAGAAGTGACGTCGGGCCATTCAGCTCGACGAAGGCT GTATAGATGCG	55
L3_c5_temp[b]	AGTCACACAGAAGTGACGTCGAGGTTAACATGGGCACGAAGGCT GTATAGATGCG	55
L2'_c1_temp[b]	GACATCATGAGAGACATCGCGTGTTACGTTAAACCGAGATACAGT GAGCGTCATG	55
L2'_c2_temp[b]	GACATCATGAGAGACATCGCTCTCTTCACACACAAGAGATACAGT GAG CGTCATG	55
L2'_c3_temp[b]	GACATCATGAGAGACATCGCGTCTTACGTTAAAGCGAGATACAGT GAGCGTCATG	55
L2'_c4_temp[b]	GACATCATGAGAGACATCGCCACCTTGAGTCTAACGAGATACAGT GAGCGTCATG	55
L2'_c5_temp[b]	GACATCATGAGAGACATCGCTCGTTTGAGGCTAACGAGATACAGT GAGCGTCATG	55
L2'_c6_temp[b]	GACATCATGAGAGACATCGCTGTTACGTTAAACGCGAGATACAGT GAGCGTCATG	55
L2'_c7_temp[b]	GACATCATGAGAGACATCGCTACCTTGAGTCTAACGAGATACAGT GAGCGTCATG	55
L2'_c8_temp[b]	GACATCATGAGAGACATCGCTCTTACGTTAAAGGCGAGATACAGT GAGCGTCATG	55
L2'_c9_temp[b]	GACATCATGAGAGACATCGCTACCTTGAGGCTTACGAGATACAGT GAGCGTCATG	55
L2'_c10_temp[b]	GACATCATGAGAGACATCGCTGTTACGTTAAACCCGAGATACAGT GAGCGTCATG	55
L2'_c11_temp[b]	GACATCATGAGAGACATCGCTACCTTGAGGCTAACGAGATACAGT GAGCGTCATG	55
L2'_c12_temp[b]	GACATCATGAGAGACATCGCACCCTTGAGTCTAACGAGATACAGT GAGCGTCATG	55
L2'_c13_temp[b]	GACATCATGAGAGACATCGCCTGTTACGTTAAACGGAGATACAGT GAGCGTCATG	55
L2'_c14_temp[b]	GACATCATGAGAGACATCGCACGTTTGAGGCTAACGAGATACAGT GAGCGTCATG	55
L2'_c15_temp[b]	GACATCATGAGAGACATCGCGCTCGATCTACAAAAGAGATACAGT GAGCGTCATG	55
L2'_c16_temp[b]	GACATCATGAGAGACATCGCTCGGGATCTACAAAAGAGATACAGT GAGCGTCATG	55
L2'_c17_temp[b]	GACATCATGAGAGACATCGCGCATATCATGATAACGAGATACAGT GAGCGTCATG	55
L2'_c18_temp[b]	GACATCATGAGAGACATCGCTCCGTTGAGTCTAACGAGATACAGT GAGCGTCATG	55
L2'_c19_temp[b]	GACATCATGAGAGACATCGCTTGTAACAATCGCAAGAGATACAGT GAGCGTCATG	55
L2'_c20_temp[b]	GACATCATGAGAGACATCGCTTGTAACAATCGGCAGAGATACAGT GAGCGTCATG	55
L2'_c21_temp[b]	GACATCATGAGAGACATCGCCCATATCATGATAACGAGATACAGT GAG CGTCATG	55

L2'_c22_temp[b]	GACATCATGAGAGACATCGCTTTCATAAGCGACATGAGATACAGT GAG CGTCATG	55
L2'_c23_temp[b]	GACATCATGAGAGACATCGCAACCTTGAGGCTTACGAGATACAGT GAGCGTCATG	55
L2'_c24_temp[b]	GACATCATGAGAGACATCGCTCGAGTATAATGCAAGAGATACAGT GAGCGTCATG	55
L2'_c25_temp[b]	GACATCATGAGAGACATCGCCTCTTACGTTAAAGGGAGATACAGT GAGCGTCATG	55
L2'_c26_temp[b]	GACATCATGAGAGACATCGCACCTTTGAGTCTAACGAGATACAGT GAGCGTCATG	55
L2'_c27_temp[b]	GACATCATGAGAGACATCGCGCTGGATCTACAAAAGAGATACAGT GAGCGTCATG	55
L2'_c28_temp[b]	GACATCATGAGAGACATCGCCACCTTGAGGCTTACGAGATACAGT GAGCGTCATG	55
L2'_c29_temp[b]	GACATCATGAGAGACATCGCGTTCGATCTACAAAAGAGATACAGT GAGCGTCATG	55
HIR-6_N	CATGACGCTCACTGTATCTCGCGTTTAACGTAACAGCGATGTCTCT CATGATGTC	55
HIR-8_N	CATGACGCTCACTGTATCTCGCCTTTAACGTAAGAGCGATGTCTCT CATGATGTC	55
SC_temp[b]	CTACGACTCAATGGTAGTATTCAACGCGTGCGAAAAGTTAAAGGC CAGGTGCAGT	55
Rev2_SC[a]	ACTGCACCTGGCCTTTAACT	20
T1_temp[b]	5CATGAGAGACATCGCTGTTACGTTAAACGCGAGATACAGTGAGC GTCATG	50
T2_temp[b]	GAGACATCGCTGTTACGTTAAACGCGAGATACAGTGAGCGTCATG	45
T3_temp[b]	ATCGCTGTTACGTTAAACGCGAGATACAGTGAGCGTCATG	40
Rev2A[a]	CGCTCACTGTATCTC	15
Rev2B[a]	ACTGTATCTC	10
Rev2C[a]	TATCTCGCG	9
Rev2D[a]	CATGACGCTCACTGTATCTCGCGTTTAA	28

[a] 5'-Cy5; [b] 5'-Biotin; [c] 5'-Phosphate; [d] mutagenized (21%) part underlined.

1.2 Libraries and oligonucleotides synthesized in this study

Table S2. Sequences of modified libraries, candidates and aptamers synthesized in this study.

ON title	Sequence (5'→3')	nt	dN^{R}TPs	Template	Primer
L1[a]	CGCATCTATACAGCCTTCGT-N15-GACGTCACTTCTGTGTGACT	55	-,-	L1_temp[b]	Rev1[a]
L2[a]	CGCATCTATACAGCCTTCGT-N15-GACGUCACUUCUGUGUGACU	55	1,2	L1_temp[b]	Rev1[a]
L3[a]	CGCATCTATACAGCCTTCGT-N15-GACGUCACUUCUGUGUGACU	55	3,4	L1_temp[b]	Rev1[a]
L2'[a]	CATGACGCTCACTGTATCTC-N15-GCGAUGUCUCUCAUGAUGUC	55	1,2	L2_temp[b]	Rev2[a]
L4[a,c]	GCTAATCTCATCAGCGTCTCCAUGACGCUCACUGUAUCUCGCGUUUAACGUAACAGCGAUGUCUCUCAUGAUGUCUGCUCGUCAGAUGUGUCUAC	95	1,2	L3_temp[b]	Rev3[a]
L2_c1[a]	CGCATCTATACAGCCTTCGTUCGACCUAUCUCGAAGACGUCACUUCUGUGUGACU	55	1,2	L2_c1_temp[b]	Rev1[a]
L2_c2[a]	CGCATCTATACAGCCTTCGTCCUUACAUCUAAUGGGACGUCACUUCUGUGUGACU	55	1,2	L2_c2_temp[b]	Rev1[a]
L2_c3[a]	CGCATCTATACAGCCTTCGTCAGACUUCGUUUGCAGACGUCACUUCUGUGUGACU	55	1,2	L2_c3_temp[b]	Rev1[a]
L2_c4[a]	CGCATCTATACAGCCTTCGTUAGACAUCUACGGGCGACGUCACUUCUGUGUGACU	55	1,2	L2_c4_temp[b]	Rev1[a]
L2_c5[a]	CGCATCTATACAGCCTTCGTGCCGCUUUAACGUAAGACGUCACUUCUGUGUGACU	55	1,2	L2_c5_temp[b]	Rev1[a]
L3_c1[a]	CGCATCTATACAGCCTTCGTGCAUCUACACUUCGCGACGUCACUUCUGUGUGACU	55	3,4	L3_c1_temp[b]	Rev1[a]
L3_c2[a]	CGCATCTATACAGCCTTCGTGCCCCCGAUAUUCGAGACGUCACUUCUGUGUGACU	55	3,4	L3_c2_temp[b]	Rev1[a]
L3_c3[a]	CGCATCTATACAGCCTTCGTAGGGUCCAUUCUAGGGACGUCACUUCUGUGUGACU	55	3,4	L3_c3_temp[b]	Rev1[a]
L3_c4[a]	CGCATCTATACAGCCTTCGTCGAGCUGAAUGGCCCGACGUCACUUCUGUGUGACU	55	3,4	L3_c4_temp[b]	Rev1[a]

L3_c5[a]	CGCATCTATACAGCCTTCGTG CCCAUGUUAACCUCGACGUC ACUUCUGUGUGACU	55	3,4	L3_c5_temp[b]	Rev1[a]
HIR-1[a]	CATGACGCTCACTGTATCTCG GUUUAACGUAACACGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c1_temp[b]	Rev2[a]
HIR-2[a]	CATGACGCTCACTGTATCTCU UGUGUGUGAAGAGAGCGAU GUCUCUCAUGAUGUC	55	1,2	L2'_c2_temp[b]	Rev2[a]
HIR-3[a]	CATGACGCTCACTGTATCTCG CUUUAACGUAAGACGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c3_temp[b]	Rev2[a]
HIR-4[a]	CATGACGCTCACTGTATCTCG UUAGACUCAAGGUGGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c4_temp[b]	Rev2[a]
HIR-5[a]	CATGACGCTCACTGTATCTCG UUAGCCUCAAACGAGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c5_temp[b]	Rev2[a]
HIR-6[a,b]	CATGACGCTCACTGTATCTCG CGUUUAACGUAACAGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c6_temp[b]	Rev2[a]
HIR-7[a]	CATGACGCTCACTGTATCTCG UUAGACUCAAGGUAGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c7_temp[b]	Rev2[a]
HIR-8[a]	CATGACGCTCACTGTATCTCG CCUUUAACGUAAGAGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c8_temp[b]	Rev2[a]
HIR-9[a]	CATGACGCTCACTGTATCTCG UAAGCCUCAAGGUAGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c9_temp[b]	Rev2[a]
HIR-10[a]	CATGACGCTCACTGTATCTCG GGUUUAACGUAACAGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c10_temp[b]	Rev2[a]
HIR-11[a]	CATGACGCTCACTGTATCTCG UUAGCCUCAAGGUAGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c11_temp[b]	Rev2[a]
HIR-12[a]	CATGACGCTCACTGTATCTCG UUAGACUCAAGGGUGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c12_temp[b]	Rev2[a]
HIR-13[a]	CATGACGCTCACTGTATCTCC GUUUAACGUAACAGGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c13_temp[b]	Rev2[a]
HIR-14[a]	CATGACGCTCACTGTATCTCG UUAGCCUCAAACGUGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c14_temp[b]	Rev2[a]
HIR-15[a]	CATGACGCTCACTGTATCTCU UUUGUAGAUCGAGCGCGAU GUCUCUCAUGAUGUC	55	1,2	L2'_c15_temp[b]	Rev2[a]

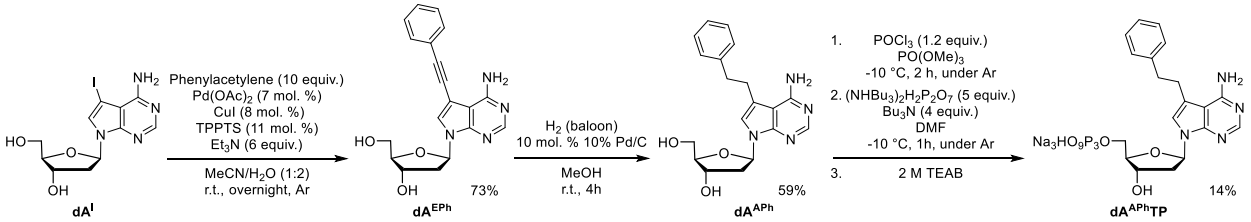
HIR-16[a]	CATGACGCTCACTGTATCTCU UUUGUAGAUCCCGAGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c16_temp[b]	Rev2[a]
HIR-17[a]	CATGACGCTCACTGTATCTCG UUAUCAUGAUAUGCGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c17_temp[b]	Rev2[a]
HIR-18[a]	CATGACGCTCACTGTATCTCG UUAGACUCAACGGAGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c18_temp[b]	Rev2[a]
HIR-19[a]	CATGACGCTCACTGTATCTCU UGCGAUUGUUACAAGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c19_temp[b]	Rev2[a]
HIR-20[a]	CATGACGCTCACTGTATCTCU GCCGAUUGUUACAAGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c20_temp[b]	Rev2[a]
HIR-21[a]	CATGACGCTCACTGTATCTCG UUAUCAUGAUAUGGGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c21_temp[b]	Rev2[a]
HIR-22[a]	CATGACGCTCACTGTATCTCA UGUCGCUUAUGAAAGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c22_temp[b]	Rev2[a]
HIR-23[a]	CATGACGCTCACTGTATCTCG UAAGCCUCAAGGUUGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c23_temp[b]	Rev2[a]
HIR-24[a]	CATGACGCTCACTGTATCTCT GCAUUAUACUCGAGCGAUGU CUCUCAUGAUGUC	55	1,2	L2'_c24_temp[b]	Rev2[a]
HIR-25[a]	CATGACGCTCACTGTATCTCC CUUUAACGUAAGAGGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c25_temp[b]	Rev2[a]
HIR-26[a]	CATGACGCTCACTGTATCTCG UUAGACUCAAAGGUGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c26_temp[b]	Rev2[a]
HIR-27[a]	CATGACGCTCACTGTATCTCU UUUGUAGAUCCAGCGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c27_temp[b]	Rev2[a]
HIR-28[a]	CATGACGCTCACTGTATCTCG UAAGCCUCAAGGUGGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c28_temp[b]	Rev2[a]
HIR-29[a]	CATGACGCTCACTGTATCTCU UUUGUAGAUCGAACGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c29_temp[b]	Rev2[a]
HIR_SC[a]	ACTGCACCTGGCCTTTAACTU UUCGCACGCGUUGAAUACUA CCAUUGAGUCGUAG	55	1,2	SC_temp[b]	Rev2_SC[a]
HIR-6_V1[a]	CATGACGCTCACTGTATCTCG CGUUUAACGUAACAGCGAUG UCUCUCAUGAUGUC	55	3,4	L2'_c6_temp[b]	Rev2[a]

HIR-6_V2[a]	CATGACGCTCACTGTATCTCG CGUUUAACGUAACAGCGAUG UCUCUCAUGAUGUC	55	3,2	L2'_c6_temp[b]	Rev2[a]
HIR-6_V3[a]	CATGACGCTCACTGTATCTCG CGUUUAACGUAACAGCGAUG UCUCUCAUGAUGUC	55	1,4	L2'_c6_temp[b]	Rev2[a]
HIR-6_V4[a]	CATGACGCTCACTGTATCTCG CGUUUAACGUAACAGCGAUG UCUCUCAUGAUGUC	55	5,2	L2'_c6_temp[b]	Rev2[a]
HIR-6_T1[a]	CATGACGCTCACTGTATCTCG CGUUUAACGUAACAGCGAUG UCUCUCAUG	50	1,2	T1_temp[b]	Rev2[a]
HIR-6_T2[a]	CATGACGCTCACTGTATCTCG CGUUUAACGUAACAGCGAUG UCUC	45	1,2	T2_temp[b]	Rev2[a]
HIR-6_T3[a]	CATGACGCTCACTGTATCTCG CGUUUAACGUAACAGCGAU	40	1,2	T3_temp[b]	Rev2[a]
HIR-6_T4[a]	CGCTCACTGTATCTCGCGUU UAACGUAACAGCGAUGUCUC UCAUGAUGUC	50	1,2	L2'_c6_temp[b]	Rev2A[a]
HIR-6_T5[a]	ACTGTATCTCGCGUUUAACG UAACAGCGAUGUCUCUCAUG AUGUC	45	1,2	L2'_c6_temp[b]	Rev2B[a]
HIR-6_T6[a]	TATCTCGCGUUUAACGUAAC AGCGAUGUCUCUCAUGAUGU C	41	1,2	L2'_c6_temp[b]	Rev2C[a]
HIR-6_T7[a]	CGCTCACTGTATCTCGCGUU UAACGUAACAGCGAU	35	1,2	T3_temp[b]	Rev2A[a]
HIR-6_T8[a]	ACTGTATCTCGCGUUUAACG UAACAGCGAU	30	1,2	T3_temp[b]	Rev2B[a]
HIR-6_T9[a]	CGCTCACTGTATCTCGCGUU UAACGUAACAGCGAUGUCUC	40	1,2	T2_temp[b]	Rev2A[a]
HIR-6_M1[a]	CATGACGCTCACTGTATCTCG CGTTTAACGUAACAGCGAUG UCUCUCAUGAUGUC	55	1,2	L2'_c6_temp[b]	Rev2D[a]
HIR-6_M2[a]	CATGACGCTCACTGTATCTCG CGUUUAACGUAACAGCGAUG TCTCTCATGATGTC	55	1,2	L2'_c6_temp[b]	**HIR-6_T3**[a]

1 = **dA^{AIn}TP**; 2 = **dU^{EPh}TP**; 3 = **dA^{EIn}TP**; 4 = **dU^{APh}TP**; 5 = **dA^{APh}TP**; [a] 5'-Cy5; [b] 5'-Biotin; [c]mutagenized (21%) part underlined; modified nucleotide positions highlighted in red.

2. Chemical synthesis and compound characterizations

All nucleobase-modified dNTPs used in this study were synthesized according to published protocols.[1]

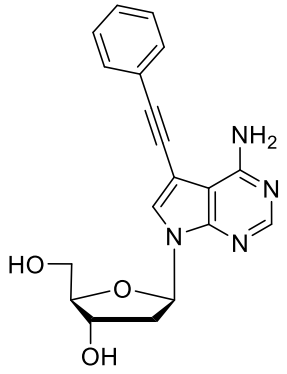


Scheme S1: Reaction pathway for synthesizing phenyl-modified **dA^{APh}TP**.

2.1 Sonogashira cross-coupling reaction

1:2 mixture of MeCN/H_2O was added through a septum to an argon-purged flask containing dA^{I} (1 equiv.), TPPTS (11 mol%), CuI (8 mol%), and $Pd(OAc)_2$ (7 mol%), followed by the addition of Phenylacetylene (10 equiv.) and TEA (6 equiv.) (Scheme S1). The reaction mixture was stirred at r.t. overnight and then evaporated under vacuum. The product was purified by FLC chromatography using DCM/MeOH (0-30%) as the eluent, followed by evaporation under vacuum to obtain a brown solid with a 73% yield.

7-(2-Phenyl-1-ethyn-1-yl)-2´-deoxyadenosine (**dA^{EPh}**)



^{1}H NMR (401 MHz, DMSO-d_6): 2.21 (ddd, 1H, J_{gem} = 13.1, $J_{2'a,1'}$ = 6.0, $J_{2'a,3'}$ = 2.8, H-2'a); 2.50 (ddd, 1H, J_{gem} = 13.1, $J_{2'b,1'}$ = 8.1, $J_{2'b,3'}$ = 5.6, H-2'b); 3.53, 3.60 (2 × ddd, 2 × 1H, J_{gem} = 11.7, $J_{5',OH}$ = 5.6, $J_{5',4'}$ = 4.4, H-5'); 3.84 (td, 1H, $J_{4',5'}$ = 4.4, $J_{4',3'}$ = 2.5, H-4'); 4.36 (dddd, 1H,

$J_{3',2'}$ = 5.6, 2.8, $J_{3',OH}$ = 4.1, $J_{3',4'}$ = 2.5, H-3'); 5.08 (t, 1H, $J_{OH,5'}$ = 5.6, OH-5'); 5.29 (d, 1H, $J_{OH,3'}$ = 4.1, OH-3'); 6.52 (dd, 1H, $J_{1',2'}$ = 8.1, 6.0, H-1'); 6.72 (bs, 2H, NH_2); 7.39 – 7.45 (m, 3H, H-*m,p*-Ph); 7.56 – 7.60 (m, 2H, H-*o*-Ph); 7.89 (s, 1H, H-6); 8.16 (s, 1H, H-2).

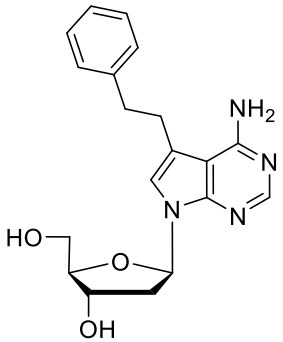
^{13}C NMR (101 MHz, DMSO-d_6): 39.92 (CH_2-2'); 61.88 (CH_2-5'); 70.96 (CH-3'); 83.06 (deazaA-**C**≡C-Ph); 83.22 (CH-1'); 87.58 (CH-4'); 91.09 (deazaA-C≡**C**-Ph); 94.68 (C-5); 102.07 (C-4a); 122.53 (C-*i*-Ph); 126.82 (CH-6); 128.50 (CH-*p*-Ph); 128.70 (CH-*m*-Ph); 131.11 (CH-*o*-Ph); 149.42 (C-7a); 152.82 (CH-2); 157.59 (C-4).

HRMS (ESI): *m/z*: $[M+H]^+$ calcd for $C_{19}H_{19}O_3N_4$: 351.14517; found: 351.14513.

2.2 Catalytic hydrogenation

MeOH was added through a septum to an argon-purged flask containing **dA^{EPh}** (1 equiv.), 10% Pd/C (10 mol%). The flask was then vacuumed and filled with H_2 atmosphere (balloon). The reaction mixture was stirred at r.t. for 4 hours and then evaporated under vacuum. The product was purified by FLC chromatography using DCM/MeOH (0-30%) as the eluent, followed by evaporation under vacuum to yield a white solid with 59% yield.

7-(2-Phenylethyl)-2´-deoxyadenosine (**dA^{APh}**)



^{1}H NMR (401 MHz, DMSO-d_6): 2.09 (ddd, 1H, J_{gem} = 13.1, $J_{2'a,1'}$ = 5.9, $J_{2'a,3'}$ = 2.6, H-2'a); 2.42 (ddd, 1H, J_{gem} = 13.1, $J_{2'b,1'}$ = 8.4, $J_{2'b,3'}$ = 5.8, H-2'b); 2.84-2.96 (m, 2H, deazaA-CH_2C**H$_2$**-Ph); 2.99-3.14 (m, 2H, deazaA-C**H$_2$**CH_2-Ph); 3.47 (ddd, 1H, J_{gem} = 11.6, $J_{5'b,OH}$ = 6.0, $J_{5'b,4'}$ = 4.6, H-5'b); 3.54 (ddd, 1H, J_{gem} = 11.6, $J_{5'a,OH}$ = 5.2, $J_{5'a,4'}$ = 4.6, H-5'a); 3.79 (td, 1H, $J_{4',5'}$ = 4.6, $J_{4',3'}$ = 2.5, H-4'); 4.30 (m, 1H, H-3'); 5.07 (dd, 1H, $J_{OH,5'}$ = 6.0, 5.2, OH-5'); 5.23 (d, 1H, $J_{OH,3'}$ = 4.1, OH-3'); 6.47 (dd, 1H, $J_{1',2'}$ = 8.4, 5.9, H-1'); 6.57 (bs, 2H, NH_2);

7.10 (t, 1H, $J_{6,CH2}$ = 1.2, H-6); 7.18 (m, 1H, H-*p*-Ph); 7.25 – 7.31 (m, 4H, H-*o*,*m*-Ph); 8.02 (s, 1H, H-2).

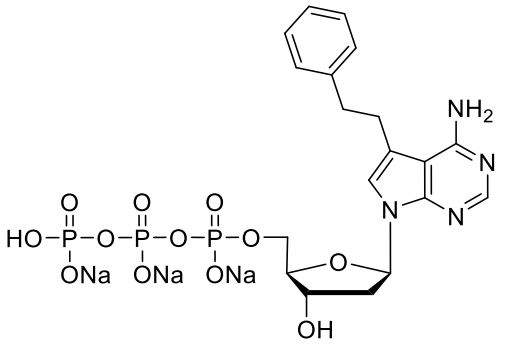
^{13}C NMR (101 MHz, DMSO-d_6): 27.61 (deazaA-**C**$H_2$$CH_2$-Ph); 35.85 (deazaA-$CH_2$**C**$H_2$-Ph); 39.51 ($CH_2$-2'); 62.17 ($CH_2$-5'); 71.12 (CH-3'); 82.73 (CH-1'); 87.11 (CH-4'); 102.20 (C-4a); 114.75 (C-5); 118.74 (CH-6); 125.78 (CH-*p*-Ph); 128.14 (CH-*m*-Ph); 128.52 (CH-*o*-Ph); 141.58 (C-*i*-Ph); 150.36 (C-7a); 151.32 (CH-2); 157.77 (C-4).

HRMS (ESI): *m/z*: $[M+H]^+$ calcd for $C_{19}H_{23}O_3N_4$: 355.17647; found: 355.17642.

2.3 Triphosphorylation

$PO(OMe)_3$ (1 mL) was added through a septum to an argon-purged flask containing **dA^{APh}** (1 equiv.), followed by the dropwise addition of $POCl_3$ (1.2 equiv.) at -10 °C (ice bath + NaCl). The reaction mixture was stirred for two hours at -10 °C. Then, the ice-cooled mixture containing a solution of $(NHBu_3)_2H_2P_2O_7$ (5 equiv.) and Bu_3N (4 equiv.) in 1 mL of dry DMF was added dropwise. The mixture was stirred for an additional hour at -10 °C. The reaction was quenched with 5 mL of aqueous 2 M TEAB (triethylammonium bicarbonate). Solvents were evaporated under vacuum and co-distilled with water three times. The product was purified using HPLC on a C18 column, employing a linear gradient from 0.1 M TEAB in water to 0.1 M TEAB in a 1:1 mixture of water and methanol as the eluent. Conversion to the sodium salt was achieved via ion exchange resin Dowex 50WX8, followed by freeze-drying from water, resulting in a white solid product with a 14% yield.

7-(2-Phenylethyl)-2´-deoxyadenosine triphosphate (**dA^{APh}TP**)



^{1}H NMR (500 MHz, D_2O): 2.36 (ddd, 1H, J_{gem} = 14.1, $J_{2'a,1'}$ = 6.3, $J_{2'a,3'}$ = 3.4, H-2'a); 2.55 (ddd, 1H, J_{gem} = 14.1, $J_{2'b,1'}$ = 8.0, $J_{2'b,3'}$ = 6.4, H-2'b); 2.92-3.04 (m, 2H, deazaA-CH_2C**H$_2$**-Ph); 3.05-3.15 (m, 2H, deazaA-C**H$_2$**CH_2-Ph); 4.05 (ddd, 1H, J_{gem} = 11.1, $J_{H,P}$ = 6.0, $J_{5'b,4'}$ = 4.5, H-5'b); 4.11 (ddd, 1H, J_{gem} = 11.1, $J_{H,P}$ = 6.6, $J_{5'a,4'}$ = 4.5, H-5'a); 4.18 (tdd, 1H, $J_{4',5'}$ = 4.5, $J_{4',3'}$ = 3.4, $J_{H,P}$ = 1.2, H-4'); 4.68 (dt, 1H, $J_{3',2'}$ = 6.4, 3.4, $J_{3',4'}$ = 3.4, H-3'); 6.57 (dd, 1H, $J_{1',2'}$ = 8.0, 6.2, H-1'); 7.08 (s, 1H, H-6); 7.14-7.18 (m, 2H, H-*o*-Ph); 7.24 (m, 1H, H-*p*-Ph); 7.27 – 7.31 (m, 2H, H-*m*-Ph); 8.08 (s, 1H, H-2).

^{13}C NMR (126 MHz, D_2O): 28.24 (deazaA-**C**H_2CH_2-Ph); 36.83 (deazaA-CH_2**C**H_2-Ph); 38.68 (CH_2-2'); 66.19 (d, $J_{C,P}$ = 5.7, CH_2-5'); 71.74 (CH-3'); 83.10 (CH-1'); 85.61 (d, $J_{C,P}$ = 8.7, CH-4'); 103.55 (C-4a); 116.82 (C-5); 120.24 (CH-6); 126.90 (CH-*p*-Ph); 129.18 (CH-*m*-Ph); 129.54 (CH-*o*-Ph); 142.12 (C-*i*-Ph); 150.51 (C-7a); 151.82 (CH-2); 158.32 (C-4).

^{31}P NMR (202 MHz, D_2O): -21.26 (dd, J = 20.4, 19.5, P_β); -10.31 (d, J = 19.5, P_α); -5.26 (d, J = 20.4, P_γ).

HRMS (ESI): *m/z*: $[M-H]^-$ calcd for $C_{19}H_{21}O_{12}N_4Na_3P_3$: 659.00674; found: 659.00662.

3. Enzymatic synthesis of DNA libraries

Doubly-modified DNA libraries (L2, L2', L3 and L4) used for selections were synthesized by primer extension (PEX) reaction in a 20 µL total volume containing 0.4 mM dNTPs (**dA^{R}TP**, **dU^{R}TP**, dCTP, dGTP, for combinations see Table S1; 0.6 mM for L4), 12.5 µM 5'-biotinylated library template (Table S1), 10 µM 5'-Cy5-labeled primer (Table S1), 1X KOD XL buffer and 2.5 U of KOD XL DNA polymerase. The reaction was incubated for 3 minutes at 95 °C, 2 minutes at 55 °C, and overnight at 60 °C. The unmodified library L1 was prepared using the same protocol but with all four natural dNTPs (dATP, dTTP, dCTP and dGTP). Twenty PEX reactions were merged together, concentrated, and purified by HPLC using a C18 column (Waters XBridge Premiere BEH Oligo 4.6x150 mm) heated to 60 °C, with a linear gradient from 0.1 M TEAB in 5% MeCN in water to 0.1 M TEAB in 20% MeCN in water over one hour, followed by a transition to 100% MeCN in one hour. Buffer pH was adjusted to 7.4 with CO_2 (g). HPLC fractions were freeze-dried, reconstituted in water, and quantified using a Nanophotometer N60 with Cy5 detection wavelength.

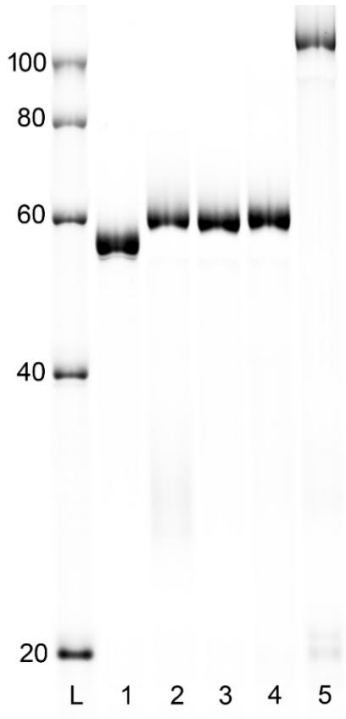


Figure S1. Cy5 scan of denaturing PAGE analysis of synthesized and HPLC-purified libraries. Cy5-labeled ssDNA ladder (lane L); L1 library (lane 1); L2 library (lane 2); L3 library (lane 3); L2' library (lane 4); L4 library (lane 5).

4. Human Insulin Receptor (HIR) as target

4.1 Amino acid sequence alignments of HIR with CIR and DIR

Sequence alignment performed with BLAST.[2]

Query = **HIR**, Accession # P06213-2; subject = **CIR**, Accession # A0A2I2UE37.

Score	Expect	Method	Identities	Positives	Gaps
2736 bits(7093)	0.0	Compositional matrix adjust.	1306/1355(96%)	1332/1355(98%)	1/1355(0%)

```
Query  1     HLYPGEVCPGMDIRNNLTRLHELENCSVIEGHLQILLMFKTRPEDFRDLSFPKLIMITDY  60
             HLYPGEVCPGMDIRNNLTRLHEL NCSVIEGHLQILLMFKTRPEDFRDLSFPKL+MITDY
Sbjct  1     HLYPGEVCPGMDIRNNLTRLHELANCSVIEGHLQILLMFKTRPEDFRDLSFPKLVMITDY  60

Query  61    LLLFRVYGLESLKDLFPNLTVIRGSRLFFNYALVIFEMVHLKELGLYNLMNITRGSVRIE  120
             LLLFRVYGLESLKDLFPNLTVIRGSRLFFNYALV+FEMVHLKELGLY+LMNITRGSVRIE
Sbjct  61    LLLFRVYGLESLKDLFPNLTVIRGSRLFFNYALVVFEMVHLKELGLYSLMNITRGSVRIE  120

Query  121   KNNELCYLATIDWSRILDSVEDNYIVLNKDDNEECGDICPGTAKGKTNCPATVINGQFVE  180
             KNNELCYLATIDWSRILDSVEDNYIVLNKDDNEECGDICPGTAKGKTNCPATVINGQFVE
Sbjct  121   KNNELCYLATIDWSRILDSVEDNYIVLNKDDNEECGDICPGTAKGKTNCPATVINGQFVE  180

Query  181   RCWTHSHCQKVCPTICKSHGCTAEGLCCHSECLGNCSQPDDPTKCVACRNFYLDGRCVET  240
             RCWTH HCQKVCPT+CKSHGCTA+GLCCHSECLGNCS+PDDPTKCVACRNFYLDGRCVET
Sbjct  181   RCWTHGHCQKVCPTVCKSHGCTADGLCCHSECLGNCSEPDDPTKCVACRNFYLDGRCVET  240

Query  241   CPPPYYHFQDWRCVNFSFCQDLHHKCKNSRRQGCHQYVIHNNKCIPECPSGYTMNSSNLL  300
             CPPPYYHFQDWRCVNFSFC+DLH+KCKNSRRQGCHQYVIHNN+CIPECPSGYTMNSSNL+
Sbjct  241   CPPPYYHFQDWRCVNFSFCRDLHNKCKNSRRQGCHQYVIHNNRCIPECPSGYTMNSSNLM  300

Query  301   CTPCLGPCPKVCHLLEGEKTIDSVTSAQELRGCTVINGSLIINIRGGNNLAAELEANLGL  360
             CTPCLGPCPKVCH+LEGEKTIDSVTSAQELRGCTVINGSLIINIRGGNNLAAELEANLGL
Sbjct  301   CTPCLGPCPKVCHILEGEKTIDSVTSAQELRGCTVINGSLIINIRGGNNLAAELEANLGL  360

Query  361   IEEISGYLKIRRSYALVSLSFFRKLRLIRGETLEIGNYSFYALDNQNLRQLWDWSKHNLT  420
             IEEISGYLKIRRSYALVSLSFFRKLRLIRGETLEIGNYSFYALDNQNLRQLWDWSKHNLT
Sbjct  361   IEEISGYLKIRRSYALVSLSFFRKLRLIRGETLEIGNYSFYALDNQNLRQLWDWSKHNLT  420

Query  421   ITQGKLFFHYNPKLCLSEIHKMEEVSGTKGRQERNDIALKTNGDQASCENELLKFSYIRT  480
             ITQGKLFFHYNPKLCLSEIHKMEEVSGTKGRQERNDIALKTNGDQASCENELLKFSYIRT
Sbjct  421   ITQGKLFFHYNPKLCLSEIHKMEEVSGTKGRQERNDIALKTNGDQASCENELLKFSYIRT  480

Query  481   SFDKILLRWEPYWPPDFRDLLGFMLFYKEAPYQNVTEFDGQDACGSNSWTVVDIDPPLRS  540
             S+DKILL+WEPYWPPDFRDLLGFMLFYKEAPYQNVTEFDGQDACGSNSWTVVDIDPPLRS
Sbjct  481   SYDKILLKWEPYWPPDFRDLLGFMLFYKEAPYQNVTEFDGQDACGSNSWTVVDIDPPLRS  540

Query  541   NDPKSQNHPGWLMRGLKPWTQYAIFVKTLVTFSDERRTYGAKSDIIYVQTDATNPSVPLD  600
             NDPKSQNHPGWLMRGLKPWTQYAIFVKTLVTFSDERRTYGAKSDIIYVQTDATNPSVPLD
Sbjct  541   NDPKSQNHPGWLMRGLKPWTQYAIFVKTLVTFSDERRTYGAKSDIIYVQTDATNPSVPLD  600

Query  601   PISVSNSSSQIILKWKPPSDPNGNITHYLVFWERQAEDSELFELDYCLKGLKLPSRTWSP  660
             PISVSNSSSQIILKWKPPSDPNGNITHYLVFWERQAEDSEL+ELDYCLKGLKLPSRTWSP
Sbjct  601   PISVSNSSSQIILKWKPPSDPNGNITHYLVFWERQAEDSELYELDYCLKGLKLPSRTWSP  660

Query  661   PFESEDSQKHNQSEYEDSAGECCSCPKTDSQILKELEESSFRKTFEDYLHNVVFVPRKTS  720
             PFESE SQK NQSEYE+SAGECCSCPKTDSQILKELEESSFRKTFEDYLHNVVFVPRK+
Sbjct  661   PFESEGSQKRNQSEYEESAGECCSCPKTDSQILKELEESSFRKTFEDYLHNVVFVPRKSF  720

Query  721   SGTGAEDPRPSRKRRSLGDVGNVTVAVPTVAAFPNTSSTSVPTSPEEHRPFEKVVNKESL  780
             S  GAED RPSRKRR+L D GNVT AVPTV  FPNTSS SVPTSPEEH+PFEKVVNKESL
Sbjct  721   SSPGAEDSRPSRKRRALADAGNVTAAVPTVPGFPNTSS-SVPTSPEEHKPFEKVVNKESL  779
```

```
Query  781   VISGLRHFTGYRIELQACNQDTPEERCSVAAYVSARTMPEAKADDIVGPVTHEIFENNVV  840
             VISGLRHFTGYRIELQACNQD PEERCSVAAYVSARTMPEAKADDIVGPVTHEIFENNVV
Sbjct  780   VISGLRHFTGYRIELQACNQDAPEERCSVAAYVSARTMPEAKADDIVGPVTHEIFENNVV  839

Query  841   HLMWQEPKEPNGLIVLYEVSYRRYGDEELHLCVSRKHFALERGCRLRGLSPGNYSVRIRA  900
             HLMWQEPKEPNGLIVLYEVSYRRYGDEELHLCVSR+HFALERGCRLRGL PGNYSVR+RA
Sbjct  840   HLMWQEPKEPNGLIVLYEVSYRRYGDEELHLCVSRRHFALERGCRLRGLPPGNYSVRVRA  899

Query  901   TSLAGNGSWTEPTYFYVTDYLDVPSNIAKIIIGPLIFVFLFSVVIGSIYLFLRKRQPDGP  960
             TSLAGNGSWTE TYFYVTDYLDVPSNIAKIIIGPLIFVFLFSVVIGSIYLFLRKRQPDGP
Sbjct  900   TSLAGNGSWTEATYFYVTDYLDVPSNIAKIIIGPLIFVFLFSVVIGSIYLFLRKRQPDGP  959

Query  961   LGPLYASSNPEYLSASDVFPCSVYVPDEWEVSREKITLLRELGQGSFGMVYEGNARDIIK  1020
             LGPLYASSNPEYLSASDVFPCSVYVPDEWEV REKITLLRELGQGSFGMVYEGNARDI+K
Sbjct  960   LGPLYASSNPEYLSASDVFPCSVYVPDEWEVPREKITLLRELGQGSFGMVYEGNARDIVK  1019

Query  1021  GEAETRVAVKTVNESASLRERIEFLNEASVMKGFTCHHVVRLLGVVSKGQPTLVVMELMA  1080
             GEAETRVAVKTVNESASLRERIEFLNEASVMKGFTCHHVVRLLGVVSKGQPTLVVMELM
Sbjct  1020  GEAETRVAVKTVNESASLRERIEFLNEASVMKGFTCHHVVRLLGVVSKGQPTLVVMELMV  1079

Query  1081  HGDLKSYLRSLRPEAENNPGRPPPTLQEMIQMAAEIADGMAYLNAKKFVHRDLAARNCMV  1140
             HGDLKSYLRSLRPEAENNPGRPPPTLQEMIQMAAEIADGMAYLNAKKFVHRDLAARNCMV
Sbjct  1080  HGDLKSYLRSLRPEAENNPGRPPPTLQEMIQMAAEIADGMAYLNAKKFVHRDLAARNCMV  1139

Query  1141  AHDFTVKIGDFGMTRDIYETDYYRKGGKGLLPVRWMAPESLKDGVFTTSSDMWSFGVVLW  1200
             AHDFTVKIGDFGMTRDIYETDYYRKGGKGLLPVRWMAPESLKDGVFTTSSDMWSFGVVLW
Sbjct  1140  AHDFTVKIGDFGMTRDIYETDYYRKGGKGLLPVRWMAPESLKDGVFTTSSDMWSFGVVLW  1199

Query  1201  EITSLAEQPYQGLSNEQVLKFVMDGGYLDQPDNCPERVTDLMRMCWQFNPKMRPTFLEIV  1260
             EITSLAEQPYQGLSNEQVLKFVMDGGYLDQPDNCPERVTDLM MCWQFNPKMRPTFLEIV
Sbjct  1200  EITSLAEQPYQGLSNEQVLKFVMDGGYLDQPDNCPERVTDLMHMCWQFNPKMRPTFLEIV  1259

Query  1261  NLLKDDLHPSFPEVSFFHSEENKAPESEELEMEFEDMENVPLDRSSHCQREEAGGRDGGS  1320
             +LLKDDLHPSFPEVSFFHSEENKAPESEELEMEFEDME+VPLDR+SH QREEAGGRDG S
Sbjct  1260  DLLKDDLHPSFPEVSFFHSEENKAPESEELEMEFEDMESVPLDRASHSQREEAGGRDGVS  1319

Query  1321  SLGFKRSYEEHIPYTHMNGGKKNGRILTLPRSNPS  1355
             SLG KR+YE+HIPYTHMNGGKKNGRILTLPRSNPS
Sbjct  1320  SLGLKRNYEDHIPYTHMNGGKKNGRILTLPRSNPS  1354
```

Figure S2. Amino acid sequence alignment of HIR and CIR. Amino acid sequences are presented in single-letter code, non-identical residues are presented with a gap, and conserved missense mutations are represented with +. Amino acid numbering starts after the signal peptide.

Query = **HIR**, Accession # P06213-2; subject = **DIR**, Accession # A0A8I3PWD4.

Score	Expect	Method	Identities	Positives	Gaps
2753 bits(7137)	0.0	Compositional matrix adjust.	1314/1355(97%)	1336/1355(98%)	1/1355(0%)

```
Query  1     HLYPGEVCPGMDIRNNLTRLHELENCSVIEGHLQILLMFKTRPEDFRDLSFPKLIMITDY  60
            HLYPGEVCPGMDIRNNLTRLHEL NCSVIEGHLQILLMFKTRPEDFRDLSFPKLIMITDY
Sbjct  1     HLYPGEVCPGMDIRNNLTRLHELANCSVIEGHLQILLMFKTRPEDFRDLSFPKLIMITDY  60

Query  61    LLLFRVYGLESLKDLFPNLTVIRGSRLFFNYALVIFEMVHLKELGLYNLMNITRGSVRIE  120
            LLLFRVYGLESLKDLFPNLTVIRGSRLFFNYALVIFEMVHLKELGLY+LMNITRGSVRIE
Sbjct  61    LLLFRVYGLESLKDLFPNLTVIRGSRLFFNYALVIFEMVHLKELGLYSLMNITRGSVRIE  120

Query  121   KNNELCYLATIDWSRILDSVEDNYIVLNKDDNEECGDICPGTAKGKTNCPATVINGQFVE  180
            KNNELCYLATIDWSRILDSVEDNYIVLNKDDNEECGDICPGTAKGKTNCPATVINGQFVE
Sbjct  121   KNNELCYLATIDWSRILDSVEDNYIVLNKDDNEECGDICPGTAKGKTNCPATVINGQFVE  180

Query  181   RCWTHSHCQKVCPTICKSHGCTAEGLCCHSECLGNCSQPDDPTKCVACRNFYLDGRCVET  240
            RCWTHSHCQKVCPTICKSHGCTAEGLCCHSECLGNCS+PDDPTKCVACRNFYLDGRCVET
Sbjct  181   RCWTHSHCQKVCPTICKSHGCTAEGLCCHSECLGNCSEPDDPTKCVACRNFYLDGRCVET  240

Query  241   CPPPYYHFQDWRCVNFSFCQDLHHKCKNSRRQGCHQYVIHNNKCIPECPSGYTMNSSNLL  300
            CPPPYYHFQDWRCVNFSFCQDLH+KCKNSRRQGCHQYVIHNNKCIPECPSGYTMNSSNL+
Sbjct  241   CPPPYYHFQDWRCVNFSFCQDLHNKCKNSRRQGCHQYVIHNNKCIPECPSGYTMNSSNLM  300

Query  301   CTPCLGPCPKVCHLLEGEKTIDSVTSAQELRGCTVINGSLIINIRGGNNLAAELEANLGL  360
            CTPCLGPCPKVCH+LEGEKTIDSVTSAQELRGCTV+NGSLIINIRGGNNLAAELEANLGL
Sbjct  301   CTPCLGPCPKVCHILEGEKTIDSVTSAQELRGCTVVNGSLIINIRGGNNLAAELEANLGL  360

Query  361   IEEISGYLKIRRSYALVSLSFFRKLRLIRGETLEIGNYSFYALDNQNLRQLWDWSKHNLT  420
            IEEISGYLKIRRSYALVSLSFFRKLRLIRGETLEIGNYSFYALDNQNLRQLWDWSKHNLT
Sbjct  361   IEEISGYLKIRRSYALVSLSFFRKLRLIRGETLEIGNYSFYALDNQNLRQLWDWSKHNLT  420

Query  421   ITQGKLFFHYNPKLCLSEIHKMEEVSGTKGRQERNDIALKTNGDQASCENELLKFSYIRT  480
            ITQGKLFFHYNPKLCLSEIHKMEEVSGTKGRQERNDIALKTNGDQASCENELLKFSYIRT
Sbjct  421   ITQGKLFFHYNPKLCLSEIHKMEEVSGTKGRQERNDIALKTNGDQASCENELLKFSYIRT  480

Query  481   SFDKILLRWEPYWPPDFRDLLGFMLFYKEAPYQNVTEFDGQDACGSNSWTVVDIDPPLRS  540
            S+DKILL+WEPYWPPDFRDLLGFMLFYKEAPYQNVTEFDGQDACGSNSWTVVDIDPPLRS
Sbjct  481   SYDKILLKWEPYWPPDFRDLLGFMLFYKEAPYQNVTEFDGQDACGSNSWTVVDIDPPLRS  540

Query  541   NDPKSQNHPGWLMRGLKPWTQYAIFVKTLVTFSDERRTYGAKSDIIYVQTDATNPSVPLD  600
            NDPKSQNHPGWLMRGLKPWTQYAIFVKTLVTFSDERRTYGAKSDIIYVQTDATNPSVPLD
Sbjct  541   NDPKSQNHPGWLMRGLKPWTQYAIFVKTLVTFSDERRTYGAKSDIIYVQTDATNPSVPLD  600

Query  601   PISVSNSSSQIILKWKPPSDPNGNITHYLVFWERQAEDSELFELDYCLKGLKLPSRTWSP  660
            PISVSNSSSQIILKWKPPSDPNGNITHYLVFWERQAEDSEL+ELDYCLKGLKLPSRTWSP
Sbjct  601   PISVSNSSSQIILKWKPPSDPNGNITHYLVFWERQAEDSELYELDYCLKGLKLPSRTWSP  660

Query  661   PFESEDSQKHNQSEYEDSAGECCSCPKTDSQILKELEESSFRKTFEDYLHNVVFVPRKTS  720
            PFE+E SQKHNQSEYE+SAGECCSCPKTDSQILKELEESSFRKTFEDYLHNVVFVPRK+S
Sbjct  661   PFEAEGSQKHNQSEYEESAGECCSCPKTDSQILKELEESSFRKTFEDYLHNVVFVPRKSS  720

Query  721   SGTGAEDPRPSRKRRSLGDVGNVTVAVPTVAAFPNTSSTSVPTSPEEHRPFEKVVNKESL  780
            S  GAED RPSRKRR+L D GNVT A+PTV  FPNT STS PTSPEEH+PFEKVVNKESL
Sbjct  721   SSPGAEDTRPSRKRRALSDAGNVTAAIPTVPGFPNT-STSTPTSPEEHKPFEKVVNKESL  779
```

```
Query  781   VISGLRHFTGYRIELQACNQDTPEERCSVAAYVSARTMPEAKADDIVGPVTHEIFENNVV  840
             VISGLRHFTGYRIELQACNQD+PEERCSVAAYVSARTMPEAKADDIVGPVTHEIFENNVV
Sbjct  780   VISGLRHFTGYRIELQACNQDSPEERCSVAAYVSARTMPEAKADDIVGPVTHEIFENNVV  839

Query  841   HLMWQEPKEPNGLIVLYEVSYRRYGDEELHLCVSRKHFALERGCRLRGLSPGNYSVRIRA  900
             HLMWQEPKEPNGLIVLYEVSYRRYGDEELHLCVSR+HFALERGCRLRGL PGNYSVR+RA
Sbjct  840   HLMWQEPKEPNGLIVLYEVSYRRYGDEELHLCVSRRHFALERGCRLRGLPPGNYSVRVRA  899

Query  901   TSLAGNGSWTEPTYFYVTDYLDVPSNIAKIIIGPLIFVFLFSVVIGSIYLFLRKRQPDGP  960
             TSLAGNGSWTE TYFYVTDYLDVPSNIAKIIIGPLIFVFLFSVVIGSIYLFLRKRQPDGP
Sbjct  900   TSLAGNGSWTEATYFYVTDYLDVPSNIAKIIIGPLIFVFLFSVVIGSIYLFLRKRQPDGP  959

Query  961   LGPLYASSNPEYLSASDVFPCSVYVPDEWEVSREKITLLRELGQGSFGMVYEGNARDIIK  1020
             LGPLYASSNPEYLSASDVFPCSVYVPDEWEV REKITLLRELGQGSFGMVYEGNARDI+K
Sbjct  960   LGPLYASSNPEYLSASDVFPCSVYVPDEWEVPREKITLLRELGQGSFGMVYEGNARDIVK  1019

Query  1021  GEAETRVAVKTVNESASLRERIEFLNEASVMKGFTCHHVVRLLGVVSKGQPTLVVMELMA  1080
             GE ETRVAVKTVNESASLRERIEFLNEASVMKGFTCHHVVRLLGVVSKGQPTLVVMELMA
Sbjct  1020  GEPETRVAVKTVNESASLRERIEFLNEASVMKGFTCHHVVRLLGVVSKGQPTLVVMELMA  1079

Query  1081  HGDLKSYLRSLRPEAENNPGRPPPTLQEMIQMAAEIADGMAYLNAKKFVHRDLAARNCMV  1140
             HGDLKSYLRSLRPEAENNPGRPPPTLQEMIQMAAEIADGMAYLNAKKFVHRDLAARNCMV
Sbjct  1080  HGDLKSYLRSLRPEAENNPGRPPPTLQEMIQMAAEIADGMAYLNAKKFVHRDLAARNCMV  1139

Query  1141  AHDFTVKIGDFGMTRDIYETDYYRKGGKGLLPVRWMAPESLKDGVFTTSSDMWSFGVVLW  1200
             AHDFTVKIGDFGMTRDIYETDYYRKGGKGLLPVRWMAPESLKDGVFTTSSDMWSFGVVLW
Sbjct  1140  AHDFTVKIGDFGMTRDIYETDYYRKGGKGLLPVRWMAPESLKDGVFTTSSDMWSFGVVLW  1199

Query  1201  EITSLAEQPYQGLSNEQVLKFVMDGGYLDQPDNCPERVTDLMRMCWQFNPKMRPTFLEIV  1260
             EITSLAEQPYQGLSNEQVLKFVMDGGYLDQPDNCPERVTDLMRMCWQFNPKMRPTFLEIV
Sbjct  1200  EITSLAEQPYQGLSNEQVLKFVMDGGYLDQPDNCPERVTDLMRMCWQFNPKMRPTFLEIV  1259

Query  1261  NLLKDDLHPSFPEVSFFHSEENKAPESEELEMEFEDMENVPLDRSSHCQREEAGGRDGGS  1320
             +LLKDDLHPSFPEVSFFHSEENKAPESEELEMEFEDME+VPLDR+SH QREEAGGRD GS
Sbjct  1260  DLLKDDLHPSFPEVSFFHSEENKAPESEELEMEFEDMESVPLDRASHSQREEAGGRDAGS  1319

Query  1321  SLGFKRSYEEHIPYTHMNGGKKNGRILTLPRSNPS  1355
             SLG KRSYEEHIPYTHMNGGKKNGRILTLPRSNPS
Sbjct  1320  SLGLKRSYEEHIPYTHMNGGKKNGRILTLPRSNPS  1354
```

Figure S3. Amino acid sequence alignment of HIR and DIR. Amino acid sequences are presented in single-letter code, non-identical residues are presented with a gap, and conserved missense mutations are represented with +. Amino acid numbering starts after the signal peptide.

4.2 Immobilization of HIR

HIR was immobilized on Dynabeads His-Tag Isolation & Pulldown beads: 10 µL of resuspended beads were collected on a magnet to remove the supernatant and incubated with His-tagged HIR (24 µg, 150% of theoretical beads capacity) and His-tagged peptide (4 µg, 25% of theoretical beads capacity, used as a surface blocker) in 500 µL of binding buffer (50 mM sodium phosphate, 300 mM NaCl, 0.01% Tween-20, pH 8) for 30 minutes at r.t. using a Hulamixer. Beads were washed twice with 400 µL of binding buffer, once with 400 µL of PBS and then collected on a

magnet. Supernatant was removed and HIR-functionalized beads were resuspended in 10 µL of PBS. The immobilized HIR was quantified with the BCA Protein Assay Kit (Novagen) following the manufacturer's protocol. Briefly, a series of BSA dilutions (20-2000 µg/mL) was used to create a linear calibration plot within the expected HIR concentration range. Beads functionalized only with His-tagged peptide, HIR-functionalized beads and BSA dilutions were mixed with BCA working reagent (1:20) and incubated for three hours at r.t. using a Hulamixer. Absorbance was measured at 562 nm with the Nanophotometer N60, and the concentration was determined from the calibration curve equation.

5. Aptamer selections

5.1 Single-round aptamer selection

Aptamer selections involved a single round with two partitioning steps (Table S3). The first step consisted of a mixture containing 300 pmol of pre-folded (incubated 3 minutes at 95 ºC and 10 minutes at r.t.) library (L1, L2, L2' or L3), 1000 nM HIR-functionalized magnetic beads, 1% HSA and 1 µM Prothrombin. This interaction mixture was incubated at r.t. for 30 minutes in interaction buffer (IB, PBS, 5 mM $MgCl_2$, pH 7.4). Non-binding sequences were removed by three washing steps (300 µL of IB for 4 minutes). Binding sequences were then thermally eluted into water at 95 ºC with shaking at 1000 rpm for 5 min. The second partitioning step consisted of a mixture containing the eluted binding sequences and 150 nM HIR-functionalized magnetic beads in IB. This mixture was incubated for 30 minutes at r.t., followed by 15 minutes of incubation with 800 µL of 3 mM $DxSO_4$ in IB. Non-binding sequences were removed by five washing steps (300 µL of IB for 4 minutes) and the binding sequences were again thermally eluted. The optimal number of PCR cycles was determined by qPCR and verified by native agarose gel electrophoresis.

Table S3. HIR single-round selection conditions.

Partition step	HIR (nM)	Library (pmol)	Protein competitors	Anionic competitor	Interaction volume (µL)
1st	1000	300	Yes	No	100
2nd	150	~$5x10^{-5}$	No	Yes	100

5.2 Optimal PCR cycle determination

Eluted sequences were amplified in a 20 µL qPCR reaction mixture containing 0.4 µM forward and reverse adapter primers (Table S1) and 1X SsoAdvance Universal SYBR Green Supermix (Bio-Rad). The reaction mixture was incubated at 95 °C for 2 minutes, followed by 30 cycles at 95 °C for 15 seconds and 60 °C for 30 seconds. The optimal amplification cycles were determined when a relatively high level of fluorescence was recorded.

5.3 Selection for covariation analysis

Covariation analysis consisted of four selection rounds. The first incubation mixture contained 200 pmol of pre-folded (3 minutes at 90 ºC and 10 minutes at r.t.) L4 library, 100 nM HIR-functionalized beads, 1% HSA and 1 µM Prothrombin. The interaction mixture was incubated at r.t. for 30 minutes in IB, followed by 15 minutes of incubation with 800 µL of 3 mM $DxSO_4$ in IB. Non-binding sequences were removed by 5 washing steps (300 µL of IB for 4 minutes). Binding sequences were thermally eluted into water at 95 ºC with shaking at 1000 rpm for 5 minutes. The optimal number of PCR cycles was determined by qPCR and verified by native agarose gel electrophoresis. During subsequent cycles, selection stringency was increased (Table S4).

Table S4. Selection conditions used in covariation analysis.

Cycle	HIR (nM)	L4 library (pmol)	Protein competitors	Anionic competitor (µL)	Interaction volume (µL)
1st	100	200	Yes	800	100
2nd	58	7	Yes	800	100
3rd	15	7	Yes	800	100
4th	7	7	Yes	1600	100

5.4 Generation of single-stranded DNA

Eluted sequences from the covariation analysis were PCR-amplified to obtain natural dsDNA. The 20 µL PCR mixture contained 5 µL of eluted sequences, 0.15 mM dNTPs, 0.4 µM 5'-biotinylated Fw3 primer (Table S1), 0.4 µM 5'-phosphorylated Rev3 primer (Table S1), 1X KOD XL buffer and 1.25 U of KOD XL DNA polymerase. The reaction mixture was incubated at 95 °C for 3 minutes,

followed by the qPCR-determined number of cycles (95 °C for 30 seconds, 60 °C for 30 seconds, and 72 °C for 30 seconds). If needed, additional PCR reactions were performed using the described conditions with 1 pmol of dsDNA as template. Products were purified using the Monarch PCR & DNA purification kit (NEB), quantified, and verified by native agarose gel electrophoresis. The purified dsDNA was then used for λ-exonuclease digestion of the 5'-phosphorylated reverse strand. A 50 µL digestion reaction consisted of 1X λ-exonuclease buffer, 10 U of λ-exonuclease, and 70 pmol of dsDNA, and was incubated at 37 °C and 550 rpm shaking for 2.5 hours. The 5'-biotinylated forward strand (which will later serve as template for PEX reactions with modified dN^RTPs) was purified with the Monarch PCR & DNA purification kit and verified by native agarose gel electrophoresis.

6. Next Generation Sequencing (NGS) and data analysis

6.1 Sample preparation and NGS

Recovered binding sequences from aptamer selections were prepared for NGS by two additional PCR reactions – adapter and index PCR. Three adapter PCR reactions (20 µL each) contained 5 µL of eluted sequences, 0.15 mM dNTPs, 0.4 µM forward and reverse adapter primers (Table SX), 1X KOD XL buffer, 1.25 U of KOD XL DNA polymerase and were amplified using the following thermal cycling protocol: 95 °C for 3 minutes followed by 10-18 cycles at 95 °C for 30 seconds, 60 °C for 30 seconds and 72 °C for 30 seconds. PCR products were purified using the Monarch PCR & DNA purification kit and used as templates for subsequent index PCR. This mixture (20 µL) contained 150 fmol of dsDNA from adapter PCR, 0.15 mM dNTPs, 2 µL of illumina i5 Index, 2 µL of illumina i7 Index, 1X KOD XL buffer, 1.25 U of KOD XL DNA polymerase and was incubated at 95 °C for 3 minutes followed by five cycles at 95 °C for 30 seconds, 60 °C for 30 seconds and 72 °C for 30 seconds. PCR products were extracted from a native 3% agarose gel, analyzed using a 4150 TapeStation (Agilent), and quantified with the NEBNext Ultra II DNA Library Prep Kit for Illumina, following the manufacturer´s protocol. Libraries were pooled to an equimolar ratio and paired-end sequenced using a NovaSeq system (Novogene, Cambridge). NGS data was processed with Cutadapt[3], fastq-join[4], NGmerge[5], Seqkit[6], FASTX-Toolkit[7] and FASTAptameR 2.0[8].

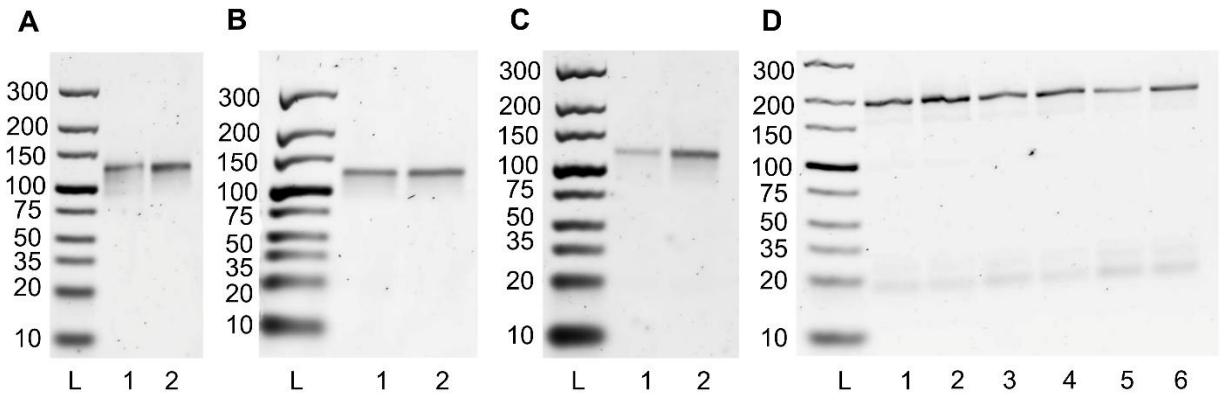


Figure S4. GelRed-stained native agarose gel of PCR-amplified sequences from SRSs. dsDNA ladder (lanes L). A), B) and C) Adapter PCR of initial libraries (lanes 1) and recovered sequences (lanes 2) from selections using L1, L2 and L3, respectively. D) Index PCR of initial libraries (odd lanes) and recovered sequences (even lanes) from selections using L1, L2 and L3, respectively.

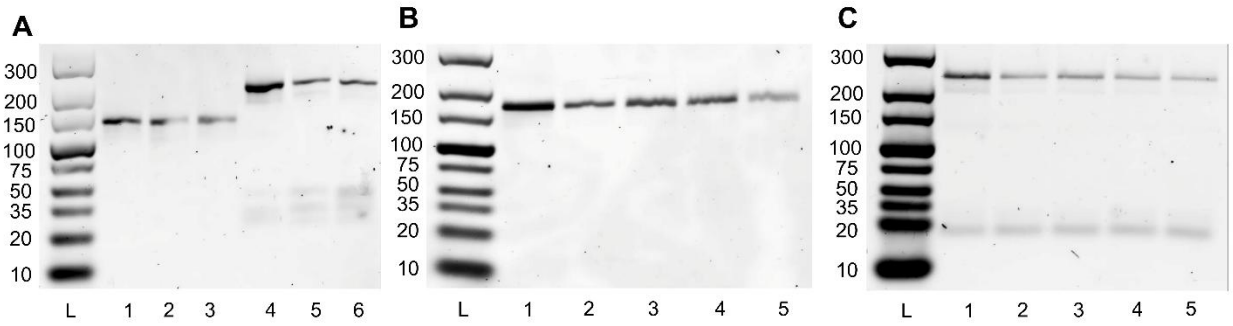


Figure S5. GelRed-stained native agarose gel of PCR-amplified sequences from SRS and covariation selection. DNA Ladder (lanes L). A) Adapter PCR (lanes 1-3) and index PCR (lanes 4-6) from selection using L2' library. Initial library (lanes 1 and 4); 1st partition step (lanes 2 and 5); 2nd partition step (lanes 3 and 6). B) Adapter and C) index PCR from covariation selection. Initial library (lanes 1); 1st round (lanes 2); 2nd round (lanes 3); 3rd round (lanes 4) and 4th round (lanes 5).

Table S5. NGS data of the SRS pools.

Selection round	Total raw sequences	Sequences used for distributions of sequence abundances
L1 - initial library	62,718,779	25,000,000
L1	61,078,238	25,000,000
L2 - initial library	53,948,409	25,000,000
L2	44,945,590	25,000,000
L3 - initial library	58,320,126	25,000,000
L3	43,383,785	25,000,000
L2' - initial library	31,439,601	28,635,776
L2' 1st partition step	32,141,588	29,407,599
L2' 2nd partition step	31,505,267	28,647,255
L4 – initial library	10,273,117	9,188,802
L4 4th round	7,471,308	6,618,962

6.2 Distributions of sequence abundances

To evaluate the impact of each library on the SRS, each sequenced pool was normalized to its initial library and used to generate enrichment profiles that display the distributions of sequence abundances. These distributions illustrate the performance of phenyl- and indol-modified dNTPs (in the case of L2 and L3) in enriching binding sequences and the evolution of the selection for L4.

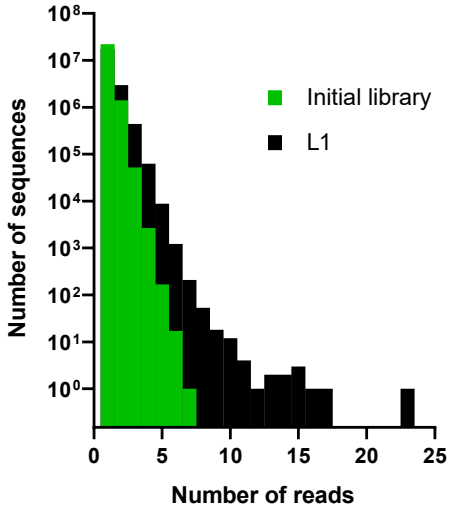


Figure S6. Distributions of sequence abundances of L1 library after SRS and its initial library.

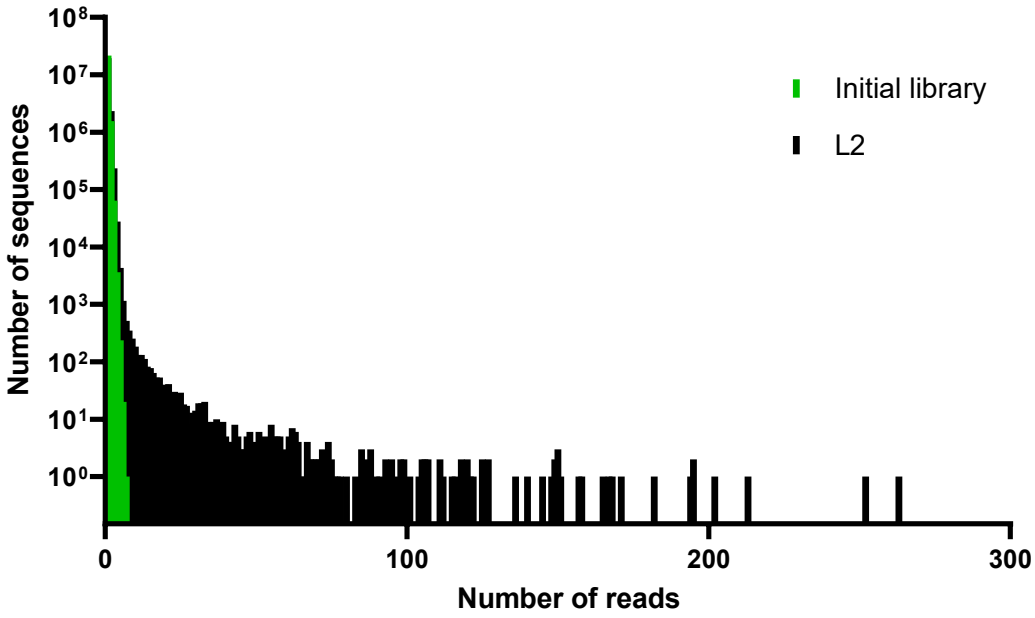


Figure S7. Distributions of sequence abundances of L2 library after SRS and its initial library.

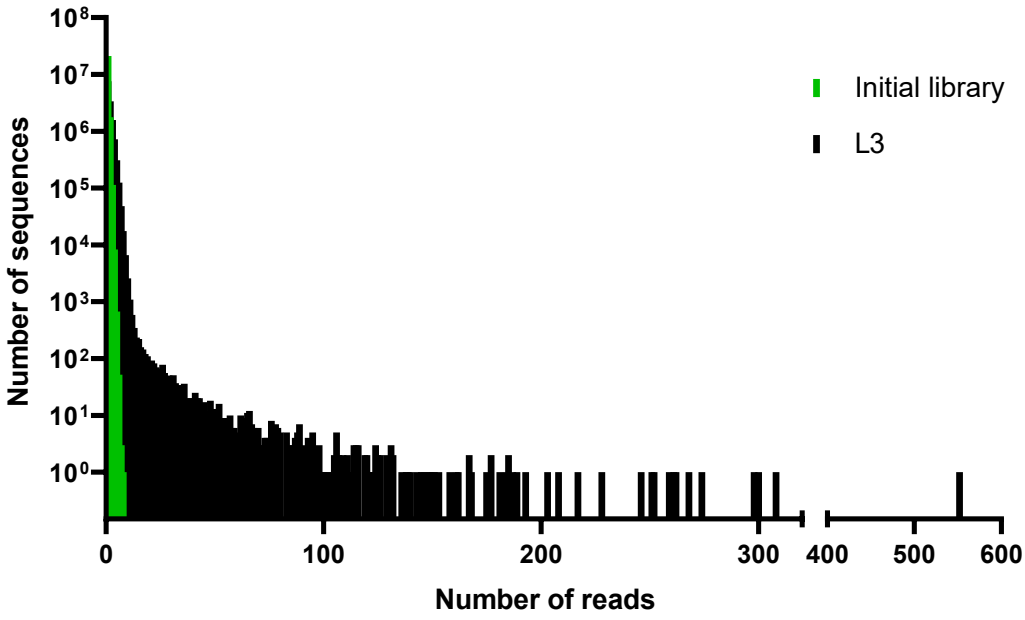


Figure S8. Distributions of sequence abundances of L3 library after SRS and its initial library.

6.3 Cluster analysis

Clusters generated from SRSs are presented in Tables S6-8. SRS using an unmodified L1 library did not produce any identifiable clusters due to the lack of enriched sequences.

Table S6. Identified clusters from SRS using L2 library.

Cluster	Seed sequence	No. sequences	Total NGS reads
1	TCGACCTATCTCGAA	29	2752
2	TAGACCTTATGTCCC	6	778
3	CGACTAGATGTTACC	4	572
4	GTGACCTATCACCGG	3	383
5	CCTTACATCTAATGG	9	761
6	CCGATTATCCCTTGG	6	762
7	CAGACTTCGTTTGCA	15	1166
8	TAGACATCTACGGGC	22	1928
9	CCATGATCTTTGAGG	6	527
10	AATTCCCAACAGGCC	3	269
11	GCCGCTTTAACGTAA	9	802
12	GTTCAGCCTAACGGT	2	223
13	GAACACCACTGTTCG	2	218
14	CGCCTATACATCTAC	2	175
15	GATCAACTGGTCACG	2	181
16	CACCTTGAGGGTTAC	3	240
17	CATCATGATATGGGC	5	420
18	TATCTGATTTGTCCC	2	211
19	GAGACCCCATAGGTG	3	236
20	ATCTCTTTGCTCACG	3	204

Table S7. Identified clusters from SRS using L3 library.

Cluster	Seed sequence	No. sequences	Total NGS reads
1	GCATCTACACTTCGC	65	7915
2	CCCTGATCAAGTCTG	13	1549
3	GCCCCCGATATTCGA	28	2496
4	CGATATTCGACCCGA	16	1387
5	AGGGTCCATTCTAGG	45	4153
6	CGAGCTGAATGGCCC	41	3489
7	GGAACTCCAATACCG	13	1261
8	GCCCATGTTAACCTC	34	2672
9	CCCCTACGTGATCTC	9	828
10	CGAACATGCTTGTCC	7	537
11	AACTCAAGTCGGGGC	27	2298

12	CCGGTGACCAACACG	8	760
13	TACGTGATCTCGACC	4	310
14	GAATTCCGTTTAGCC	1	116
15	TACGTGGTACCTGTC	2	158
16	GATCCATTAACCCCC	2	155
17	AGCTCGACATGCCGC	5	308
18	GCTGAATGGCACGGG	3	216
19	GATTGACGTACGGGC	5	309
20	ACGATTCTGTGGTGG	1	80

Table S8. Identified clusters from SRS using L2' library.

Cluster	Seed sequence	No. sequences	Total NGS reads
1	GGTTTAACGTAACAC	45	5693
2	TTGTGTGTGAAGAGA	4	609
3	GTTAGACTCAAGGTG	302	31939
4	TTTTGTAGATCGAGC	70	5915
5	GTTATCATGATATGC	19	1412
6	GTGTCGCTTATGAAA	19	1416
7	GCATTGACGCAGAAA	2	337
8	TTGACCTATCAAGAT	4	401
9	ATGATGTCCTGTCTG	2	246
10	GGTATTTGTTTACGA	4	371
11	GCGCGGTAGATGTAC	4	333
12	ATCTTGACCTATCAA	6	488
13	TTTTGTGCAATACAT	5	427
14	TTGCGATTGTTACAA	19	1328
15	TTGCATTATACTCGA	7	529
16	TGCATATATCACACT	8	619
17	TTATGCATATCCGGT	11	821
18	TACCTTGATCTTAAG	1	111
19	CGCGATGAACTGGTC	1	98
20	GCATGATGTTTTTAC	11	791

6.4 Sequence alignment of most active clusters

Figure S9. Sequence alignment of randomized regions from sequences found in cluster 1 (from L2' library) and cluster 11 (from L2 library) performed with MUSCLE[9] by MEGA12[10] and visualized with Jalview2.11.5.0[11]. Sequences with an identity of over 80% are highlighted.

6.5 Covariation analysis

Pre-processed variable regions were filtered for the correct length of 55 nucleotides, and unique sequences were counted with basic bash commands. This dataset, containing aligned sequences and their counts, was then used to calculate mutual information between all pairs of positions with a custom Python script according to the equation, and the values of mutual information were arranged into a symmetric matrix (Equation S1).

$$\mathrm{I}(X;Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log \left(\frac{P(x,y)}{P(x)P(y)} \right)$$

Equation S1. The mutual information between a pair of positions X and Y, denoted as I(X; Y), is calculated as a double sum (one over all possible values at position X and the second over all possible values at position Y) of the products of the frequencies of a given dinucleotides and the logs of ratios of frequencies of given dinucleotides (denoted as P(X,Y)(x, y)) and the products of individual frequencies of corresponding single nucleotides (denoted as PX(x) and PY(y) respectively).

7. Enzymatic synthesis of modified oligonucleotides

Aptamer candidates selected for affinity screening and subsequent truncated, mutated and scrambled sequences of **HIR-6** and **HIR-8** aptamers were prepared by PEX using a 5'-biotinylated template, 5'-Cy5-labeled reverse primer (Supplementary Table S1), and strand separated as described in the following section. If needed, more PEX reactions were performed, merged and HPLC purified. PEX conditions from Section 3.1 were applied with some variations (Table S9).

Table S9. PEX reaction conditions.

Name	dNTPs / dN^{R}TPs	KOD XL DNA polymerase (U)	Extension time (h)
HIR-6_V1/3/4	0.5 mM	3.75	3
Rest of the sequences	0.5 mM	1.25	2

7.1 Generation of single-stranded DNA using magnetic beads

The PEX reaction was strand-separated using streptavidin magnetic beads (SMB, Dynabeads MyOne Streptavidin C1) according to the manufacturer´s protocol, with some modifications: 200 µL of SMB were washed three times with 400 µL of washing buffer (5 mM Tris-HCl, 0.5 mM EDTA, 1 M NaCl, pH 7.5). PEX reaction mixture (20 µL) was mixed with 380 µL nuclease-free water and 400 µL binding buffer (10 mM Tris-HCl, 1 mM EDTA, 2 M NaCl, pH 7.5) and incubated with washed SMB for 60 minutes at r.t. using a Hulamixer. Immobilized dsDNA was washed three times with 400 µL of washing buffer and then incubated with 100 µL of 20 mM NaOH. The SMB were collected on a magnet, and the solution containing the (modified) strand was transferred into a new vial, neutralized with 0.5 M HCl, and buffered to PBS. Samples were centrifuged at 13,000×g for 3 minutes to remove any remaining SMB, quantified using a Nanophotometer N60 and analyzed by native 3% agarose (Figs. S11-14 and S17-19) and denaturing PAGE (Figs. S10, S15 and S16) gel electrophoresis.

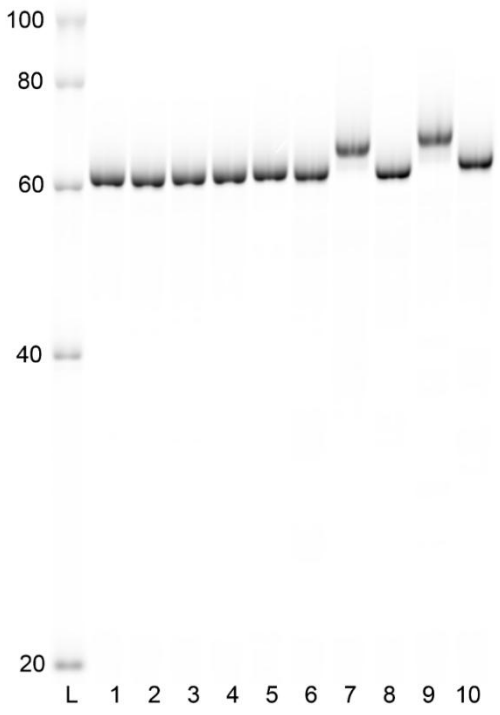


Figure S10. Cy5 scan of denaturing PAGE analysis of ten aptamer candidates from SRSs using L2 and L3 libraries. Cy5-labeled ssDNA ladder (lane L); five candidates from L2 (lanes 1-5); five candidates from L3 (lanes 6-10).

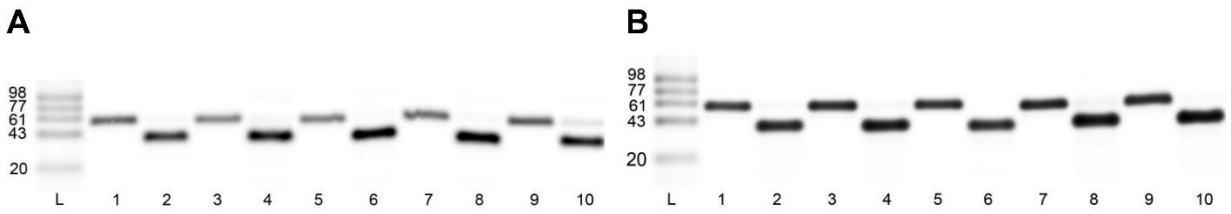


Figure S11. Cy5 scan of native agarose gel of ten aptamer candidates from SRSs using L2 and L3 libraries. Cy5-labeled ssDNA ladder (lanes L); A) five candidates from L2, dsDNA from PEX (odd lanes); modified single strands after magnetoseparation (even lanes). B) five candidates from L3, dsDNA from PEX (odd lanes); modified single strands after magnetoseparation (even lanes).

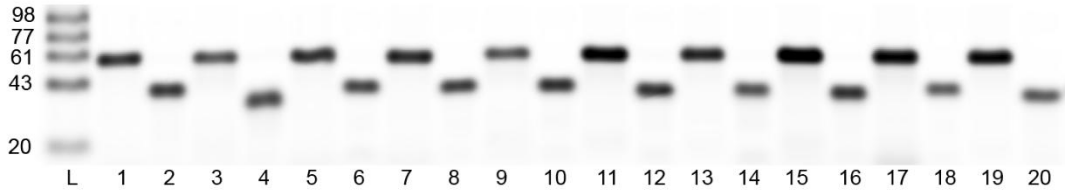


Figure S12. Cy5 scan of native agarose gel of aptamer candidates from SRS using L2' library. Cy5-labeled ssDNA ladder (lane L); **HIR-1** to **HIR-10** as dsDNA after PEX (odd lanes); **HIR-1** to **HIR-10** as ssDNA after magnetoseparation (even lanes).

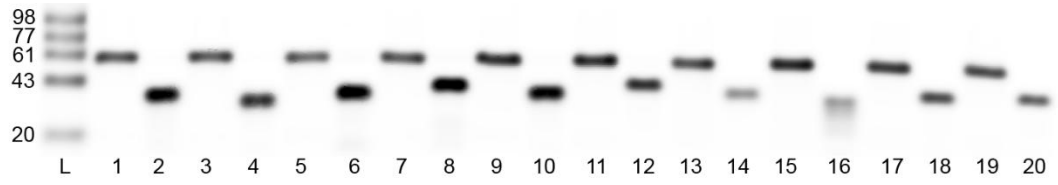


Figure S13. Cy5 scan of native agarose gel of aptamer candidates from SRS using L2' library. Cy5-labeled ssDNA ladder (lane L); **HIR-11** to **HIR-20** as dsDNA after PEX (odd lanes); **HIR-11** to **HIR-20** as ssDNA after magnetoseparation (even lanes).

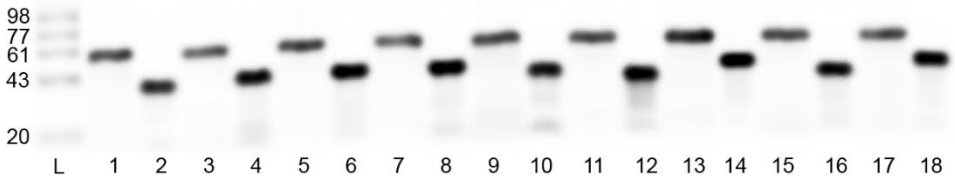


Figure S14. Cy5 scan of native agarose gel of aptamer candidates from SRS using L2' library. Cy5-labeled ssDNA ladder (lane L); **HIR-21** to **HIR-29** as dsDNA after PEX (odd lanes); **HIR-21** to **HIR-29** as ssDNA after magnetoseparation (even lanes).

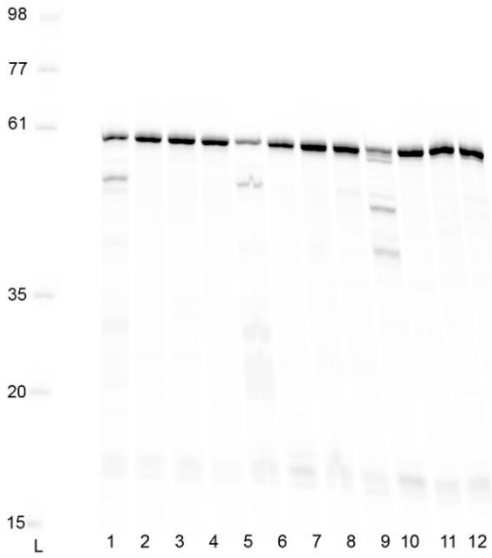


Figure S15. Cy5 scan of denaturing PAGE analysis of PEX of aptamer candidates from SRS using L2' library. Different concentrations of dN^RTPs were tested for PEX optimizations. Cy5-labeled ssDNA ladder (lane L); **HIR-1** (lanes 1-4); **HIR-2** (lanes 5-8); **HIR-3** (lanes 9-12); 0.2 mM dN^RTPs (lanes 1, 5 and 9), 0.4 mM dN^RTPs (lanes 2, 6 and 10); 0.5 mM dN^RTPs (lanes 3,7 and 11); 0.6 mM dN^RTPs (lanes 4, 8 and 12).

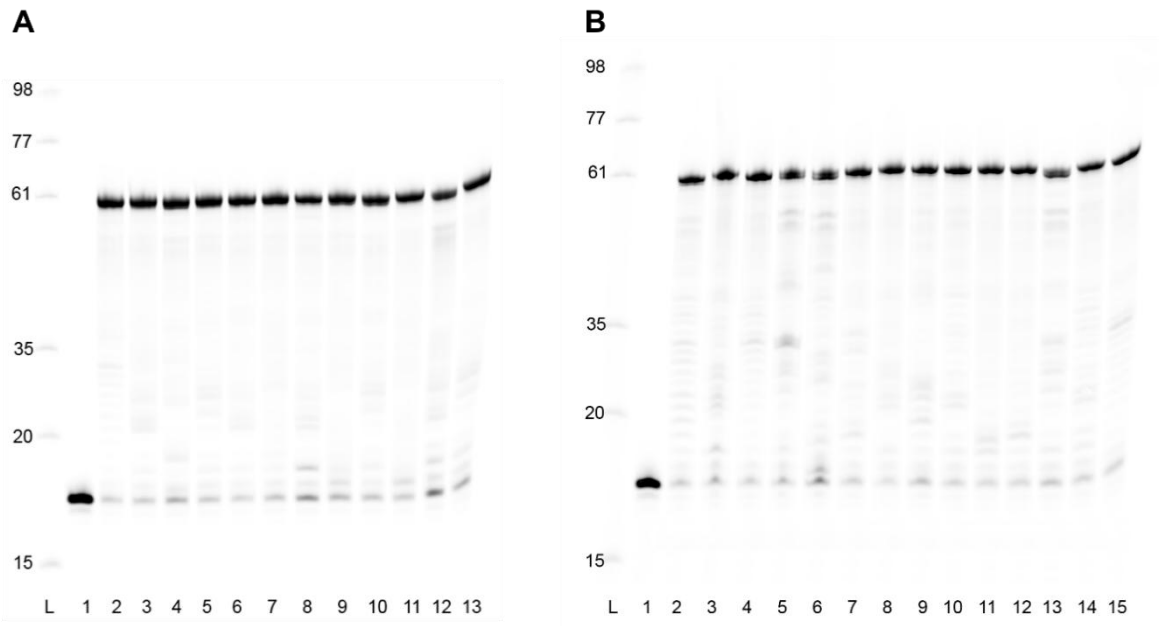


Figure S16. Cy5 scan of denaturing PAGE analysis of PEX of aptamer candidates from SRS using L2' library. Cy5-labeled ssDNA ladder (lanes L); Rev2 primer used in PEX (lanes 1); A) **HIR-4** to **HIR-15** (lanes 2-13). B) **HIR-16** to **HIR-29** (lanes 2-14).

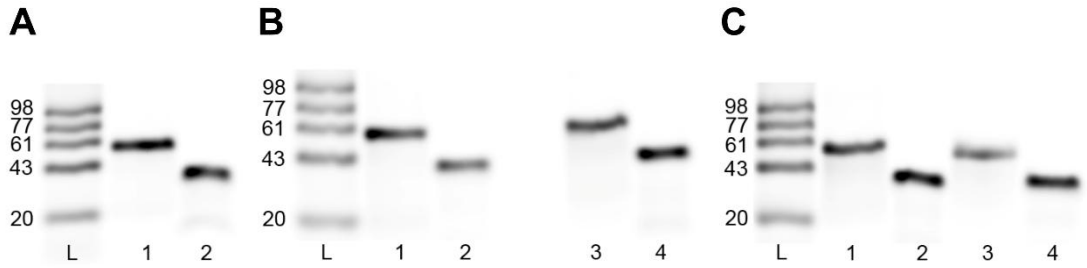


Figure S17. Cy5 scan of native agarose gel of **HIR-6** aptamer and its variant sequences. Cy5-labeled ssDNA ladder (lanes L); dsDNA after PEX (odd lanes); ssDNA after magnetoseparation (even lanes). A) **HIR-6_V2** (lanes 1-2). B) **HIR-6** (lanes 1-2); **HIR-6_V3** (lanes 3-4). C) **HIR-6_V1** (lanes 1-2); **HIR-6_V4** (lanes 3-4).

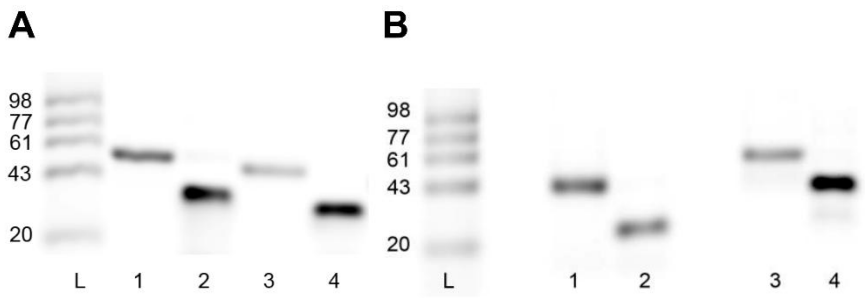


Figure S18. Cy5 scan of native agarose gel of truncated and mutated sequences of **HIR-6** aptamer. Cy5-labeled ssDNA ladder (lanes L); dsDNA after PEX (odd lanes); ssDNA after magnetoseparation (even lanes). A) **HIR-6_T1** (lanes 1-2); **HIR-6_T2** (lanes 3-4). B) **HIR-6_T3** (lanes 1-2); **HIR-6_M2** (lanes 3-4).

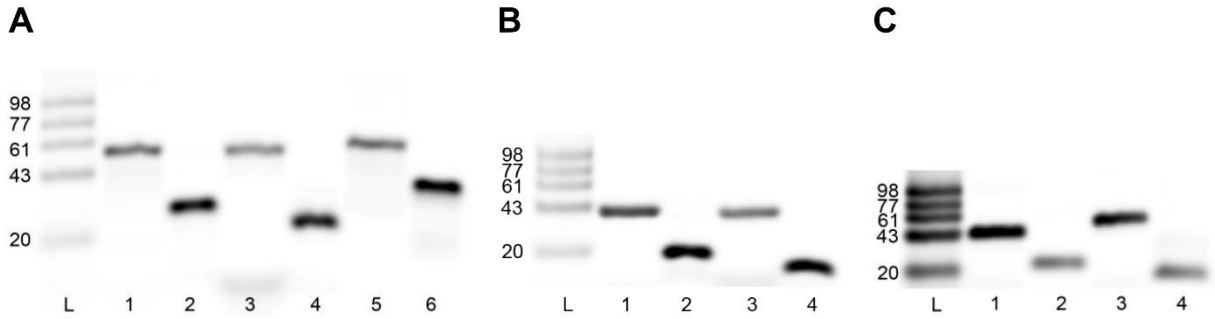


Figure S19. Cy5 scan of native agarose gel of truncated sequences of **HIR-6** aptamer. Cy5-labeled ssDNA ladder (lanes L); dsDNA after PEX (odd lanes); ssDNA after magnetoseparation (even lanes). A) **HIR-6_T4** (lanes 1-2); **HIR-6_T5** (lanes 3-4); **HIR-6_M1** (lanes 5-6). B) **HIR-6_T7** (lanes 1-2); **HIR-6_T8** (lanes 3-4). C) **HIR-6_T9** (lanes 1-2); **HIR-6_T6** (lanes 3-4).

8. Binding affinity and specificity assays

Because two identical HIR monomers form a 2-fold homodimeric symmetrical structure, each monomer can bind one aptamer molecule. Consequently, we calculated the dissociation constants (K_Ds) using a 1:1 HIR:**HIR-6** ratio using three different techniques.

8.1 Fluorescent Ni-plate binding assay

Plate assays were performed in 96-well Pierce Nickel-coated Plates. For immobilization, 100 pmol of His-tagged proteins were incubated for 2 hours at r.t. in PBS with shaking at 500 rpm. Wells were washed once with PBS containing 0.2% Tween-20 and twice with IB. Then, 100 pmol of pre-folded 5'-Cy5-labeled aptamers were incubated for 1 hour at r.t. in IB. Wells were washed three times with IB and 100 µL of IB was added for measurement. Fluorescence intensity was measured at 633ex/678em nm using a Spark multimode microplate reader (Tecan).

8.2 Microscale thermophoresis (MST)

Serial dilutions (1:1) of HIR, IGF-1 R, CIR or DIR were incubated for one hour at r.t. with 20 nM 5'-Cy5-labeled aptamers in IB containing 0.1% Pluronic F-127 to produce a protein concentration range from 210 nM to 20 pM. Samples were loaded into premium MonolithX Capillaries and measured using a MonolithX (MM-286, NanoTemper Technologies) at 25 ºC. Instrument parameters were adjusted to 100 % LED power and medium MST power. Dissociation constants (K_Ds) were calculated using the GraphPad Prism 8.01 software. Data were fitted to a specific binding with a Hill Slope non-linear regression model.

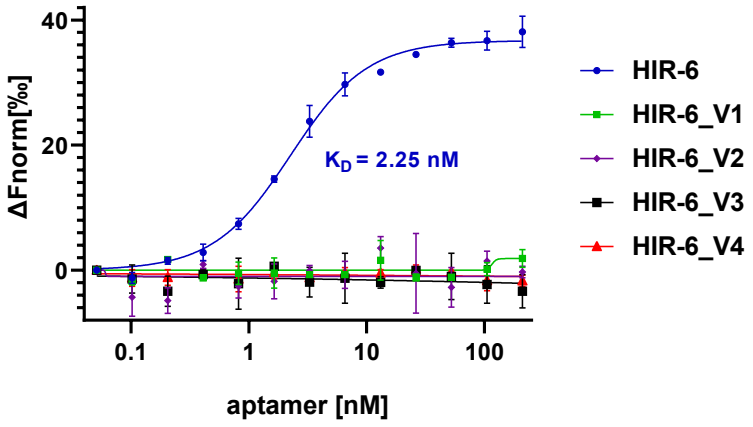


Figure S20. MST measurements of **HIR-6** and its variants with calculated K_D value. Error bars represent the mean ± s.d. from two individual experiments.

8.3 Surface plasmon resonance (SPR)

A laboratory SPR platform Plasmon IV based on the wavelength spectroscopy of surface plasmons[12] with six sensing channels and dispersionless microfluidics[13] was used. To perform **HIR-6**-HIR kinetic analysis, streptavidin was initially attached to a sensor surface using amino-coupling[14]. Then, 10 nM 5'-biotinylated **HIR-6** (pre-folded at 90 °C for 3 minutes in IB) was injected to the streptavidin-coated sensor surface for 10 minutes to reach a sensor response of ~ 0.14 nm, corresponding to approximately 2.4 ng/cm².[14] HIR concentrations ranging from 60 to 5 nM were injected for 3 minutes to the **HIR-6**-modified sensor surface (detection channel) and streptavidin-coated sensor surface (reference channel). All kinetic measurements were performed at 25 °C in an IB as running buffer at a flow rate of 20 µL/minute. A reference-compensated sensor response was obtained as a difference between the sensor response obtained in the detection and reference channel. Reference-compensated sensor responses were used for the kinetic analysis of the interactions. Kinetic and equilibrium constants were calculated assuming a 1:1 stoichiometry (1:1 Langmuir model with mass transfer) using BIAevaluation software (Biacore). Errors are reported as the standard deviation of multiple measurements.

8.4 Biolayer interferometry (BLI)

Binding kinetics were measured using an Octet RED96 (ForteBio) instrument using AR2G biosensors (Sartorius), following the manufacturer's protocol. The BLI run consisted of following steps: 60 seconds incubation with water, 300 seconds activation (400 mM EDC, 200 mM Sulfo-NHS), 600 seconds streptavidin immobilization (500 nM, acetate buffer, pH 5.5), 300 seconds quenching (1 M Ethanolamine in water), 120 seconds baseline (IB), 600 seconds loading of pre-folded 5'-biotinylated aptamer (100 nM, IB), 120 seconds baseline (IB) and 500 seconds for each association and dissociation phases. Measurements were performed at r.t. with HIR serially diluted concentrations (100 to 6.25 nM) in the association phase. One biosensor without a loaded aptamer, treated with 100 nM HIR, was used as a reference for background subtraction. Data were analyzed using GraphPad Prism 8.01 software assuming 1:1 stoichiometry.

9. Cell assays

Human IM-9 lymphocytes (ATCC, CCL-159™) and mouse embryonic fibroblasts derived from IGF-1R knockout mice stably transfected with human IR-A, kindly provided by A. Belfiore (Catanzaro, Italy) and R. Baserga (Philadelphia, Pennsylvania, USA), were grown as described previously.[15, 16, 17]

9.1 Receptor-binding studies

Binding affinities of the aptamers for IR-A were determined by the competition of compounds with [^{125}I]monoiodotyrosyl-A14-insulin for IR-A in cell membranes of human IM-9 lymphocytes (ATCC) as described previously.[15,16,17] Radiolabeled [^{125}I]monoiodotyrosyl-A14-insulin was prepared according to a published procedure.[18] The binding curve of each compound was determined in duplicate, and the final inhibition constant (K_i) was calculated as the average of two binding curves using a one-site total nonlinear regression curve with GraphPad Prism 5 software.

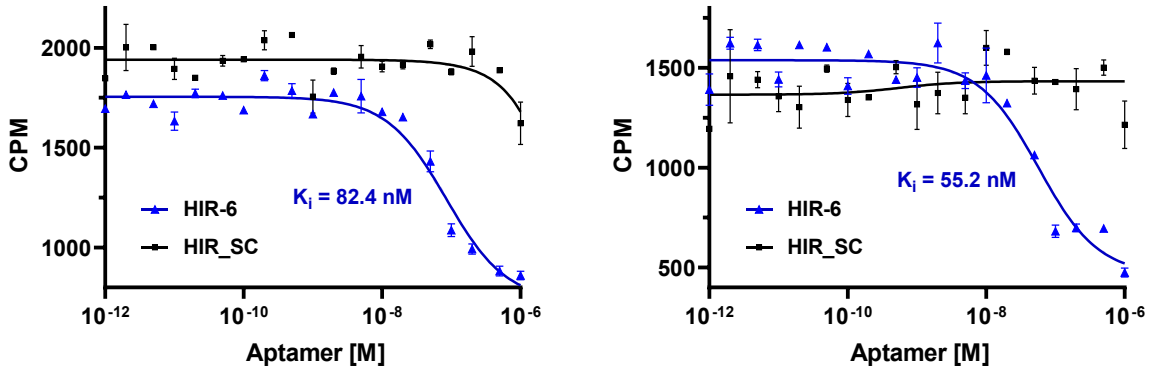


Figure S21. HIR-6 binding competition with [^{125}I]monoiodotyrosyl-A14-insulin for IR-A isoform in human IM-9 lymphocyte cell membranes using increasing concentrations of **HIR-6** and **HIR_SC**. A decrease in radioactively labeled insulin bound to HIR was measured (CPM). K_i was calculated from two technical replicates and error bars represent the mean ± s.d.

9.2 Receptor phosphorylation and antagonism assay

Ligand-dose response HIR-A autophosphorylation levels in mouse embryonic fibroblasts (IR-A) were determined using an In-Cell Western assay adapted for chemiluminescence.[19] The inhibition of insulin (at 10 nM) activity by the aptamers (concentration range from 0.1 nM to 1 μM) was performed as described.[20] Autophosphorylation of the receptor was monitored using anti-phospho-IGF-1Rβ (Tyr1135/1136)/IRβ (Tyr1150/1151) antibody (Cell Signaling Technology). Data were subtracted from the background values and expressed as the contribution of phosphorylation relative to the 10 nM insulin signal. Activation with insulin and inhibition of 10 nM insulin signaling with the aptamers was measured in duplicate. Stimulation with the aptamer was performed in mono-plicate, as no signal was observed. Control wells and wells stimulated with 10 nM insulin were conducted as tetraplicates on each plate. Each experiment was performed twice. Nonlinear regression curve fitting of the combined data from all experiments was carried out with GraphPad Prism 5 software. Control western blots (Supplementary Fig. S23) were performed according to the published protocol.[16] The IR-A cells on 24-well plates were stimulated with 500 and 100 nM **HIR-6** alone and in the presence of 10 nM insulin for 20 minutes. Proteins were routinely analyzed using immunoblotting. The membranes were cut at 75 kDa and 50 kDa standards. The appropriate parts of the membrane were probed with phospho-IGF-1Rβ (Tyr1135/1136)/IRβ (Tyr1150/1151) (19H7) (above 75 kDa), with phospho-Akt (Thr308) (C31E5E) (75-50 kDa) and Akt (pan) (C67E7) was used as a loading control. All antibodies were purchased from Cell Signaling Technology.

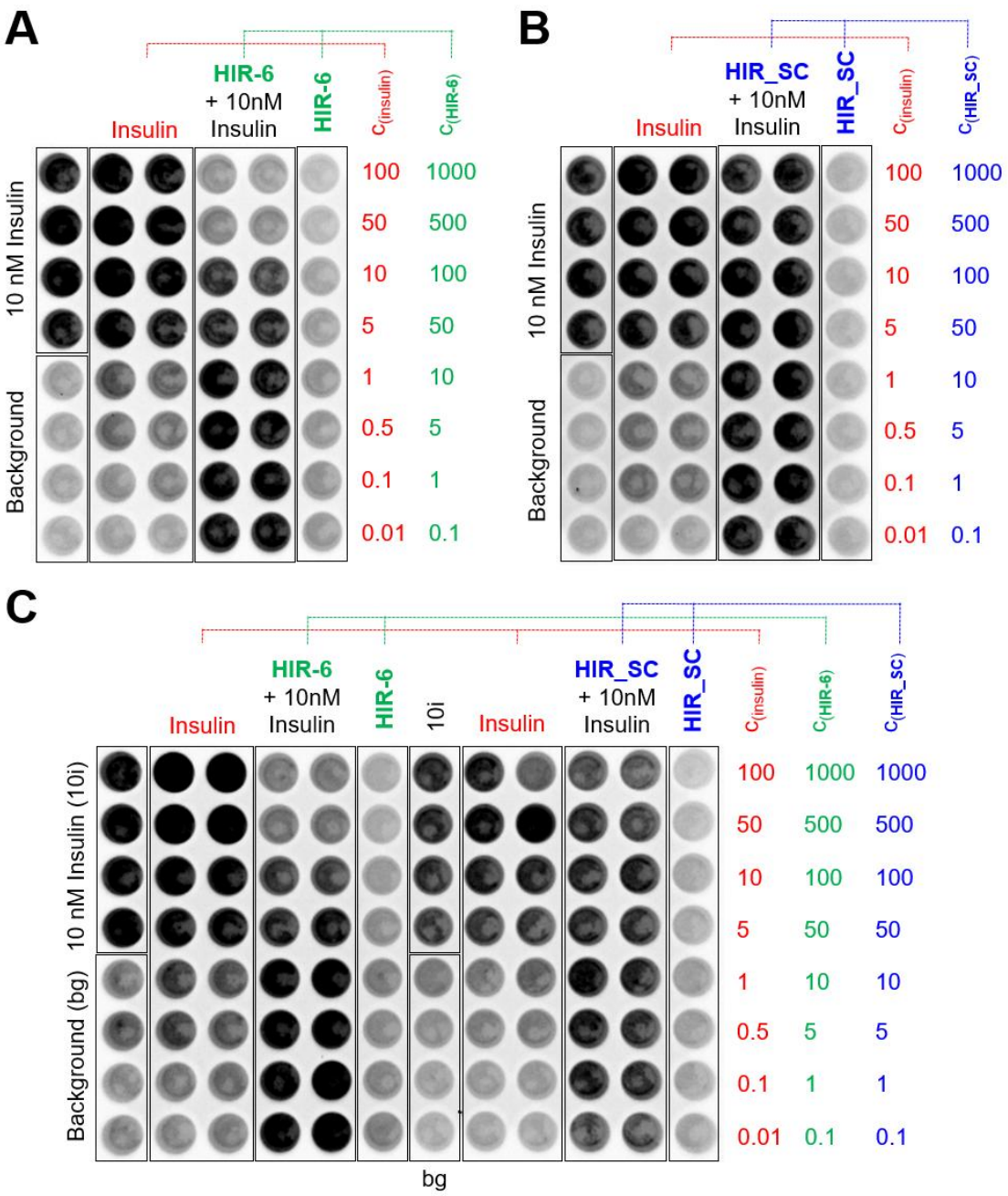


Figure S22. Stimulation of HIR phosphorylation with **HIR-6**. Individual plates from In-Cell Western Assay. The mouse fibroblasts transfected with human IR-A were treated with insulin, **HIR-6** and **HIR_SC** at specified concentrations for 20 min. Formaldehyde fixed permeabilized cells were incubated with anti-phospho-IGF-1Rβ (Tyr1135/1136)/IRβ (Tyr1150/1151) and developed with peroxidase-labeled anti-rabbit secondary antibody (Sigma). SuperSignal West Femto maximum sensitivity substrate was added to each well, and chemiluminescence was detected using the ChemiDoc MP Imaging System. Data were subtracted from the background values and expressed as the contribution of phosphorylation relative to the 10 nM insulin signal.

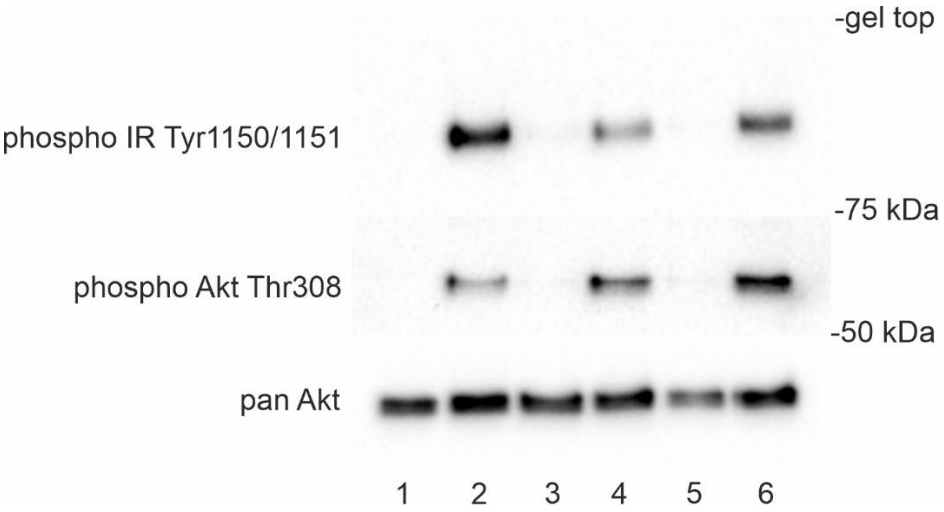


Figure S23. Control western blots showing stimulation of IR-A and Akt in mouse fibroblasts derived from IGF-1R knockout mice stably transfected with human IR-A. Cells were stimulated with 10 nM insulin (lane 2), 500 and 100 nM **HIR-6** (lanes 3 and 5, respectively). The aptamers were co-stimulated with 10 nM insulin (lane 4: 500 nM **HIR-6** and lane 6: 100 nM **HIR-6**). Control unstimulated cells are in lane 1. Membranes were cut at 75 kDa and 50 kDa standards, and respective parts were developed with anti-phospho-IGF-1Rβ (Tyr1135/1136)/IRβ (Tyr1150/1151) (Mw above 75 kDa) and anti phospho-Akt (Thr308) (C31E5E) (Mw between 75 and 50 kDa). The same samples were stained in parallel with Akt (pan) antibody used as a loading control.

10. Cryo-electron microscopy (cryo-EM) of HIR-HIR-6 complex

10.1 Cryo-EM grid preparation

HIR was reconstituted to a concentration of 3.38 μM, dialyzed to PBS using a D-tube Dialyzer mini (Merck, MWCO 6-8 kDa), and quantified with a BCA assay. The final interaction mixture (50 μL) used to prepare the cryo-EM grids contained 2 μM **HIR-6** and 0.72 μM HIR in IB. Each sample was incubated for 10 minutes at 4°C. Aliquots of 3 μL were applied to a glow-discharged Quantifoil R2/1 Au 300 mesh grid, immediately blotted for 2 seconds, and plunged into liquid ethane using a Thermo Fisher Scientific Vitrobot Mark IV (4 °C, 100% humidity).

10.2 Cryo-EM data collection

The grids were loaded into a 300 kV Titan Krios (FEI) electron microscope equipped with a Gatan K3 (model 1025) direct electron detector mounted on a Gatan BioQuantum (model 1967) energy filter. Data were collected using Serial EM software[21] in image shift acquisition mode (3 × 3 holes; 7 exposures per hole) at a nominal magnification of ×165,000 with a pixel size of 0.5113 Å per pixel. The Gatan K3 detector was used in correlated-double sampling mode and energy filter with the energy slit set to 10 eV. Movies were collected for 2 seconds at a flux of 25 electrons per $Å^2$ per second, giving a total exposure of 50 electrons per $Å^2$. The defocus values ranged from −0.5 to −3.0 µm. Forty frames of each movie were saved.

10.3 Cryo-EM image processing

All data processing (Figures S24-25, Table S10) was performed using the Relion 4.0 software package.[22] Motion correction was performed using the RELION implementation of MotionCor2.[23] Movies were aligned using 7 × 5 patches with dose weighting. A contrast transfer function (CTF) was estimated using CTFFIND4.1[24] from summed power spectra[25] for every 4 electrons per $Å^2$. 50 micrographs were randomly selected, and a representative set of particles was picked manually. These particle coordinates were used for training a Topaz picking model and Topaz model was used for subsequent particle picking.[26]

After initial binning, particles underwent two rounds of 2D-classification. In each round, the particles were sorted into 200 classes with an E-step of 8 Å and a mask diameter of 220 Å. Only classes with well-defined structural features were retained and subjected to 3D-classification using a reference from the crystal structure PDB ID 4ZXB.[27] The first 3D classification sorted particles into ten classes with the regularization parameter set to T = 4. Selected classes were aligned into a global 3D auto-refinement. A subsequent 3D classification, using the result of the previous 3D auto-refinement as input along with mask focused on the area around L2 and FnIII-1 domains and density for the **HIR-6** aptamers, was performed using local searches from 3.7 to 1.8 degrees with the regularization parameter T = 4. Particles with poorly defined structural features were removed, and the remaining particles were re-extracted to two times binned pixel size. Particles were 3D refined with focused mask on the area around L2 and FnIII-1 domains and density for the **HIR-6** aptamers, and using C2 relaxed symmetry. The 3D refinement was improved by correction for microscope aberrations within Relion 4.0. Final focused 3D

classification was performed without angular alignment and regularization parameter set to T = 20. For the final 3D refinement, the particles were Bayesian-polished and further corrected for microscope aberrations. The final cryo-EM density maps were generated by the post-processing feature in RELION and sharpened or blurred into MTZ format using CCP-EM.[28] The final set was B = -200, -100, -50, 0, 50, 100, 200 Å^2 sharpened/blurred MTZ maps. The resolution of the cryo-EM density maps (Table S10, Figures S24-25) was estimated with the gold standard Fourier shell correlation (FSC) cut-off value of 0.143. Reference-based local amplitude scaling was performed by LocScale[29] around the modelled domains of FnIII-1, L2' and **HIR-6**. The angular orientation distribution of the 3D reconstruction was calculated by cryoEF v1.1.06.[30] Local resolution was calculated within Relion 4.0.

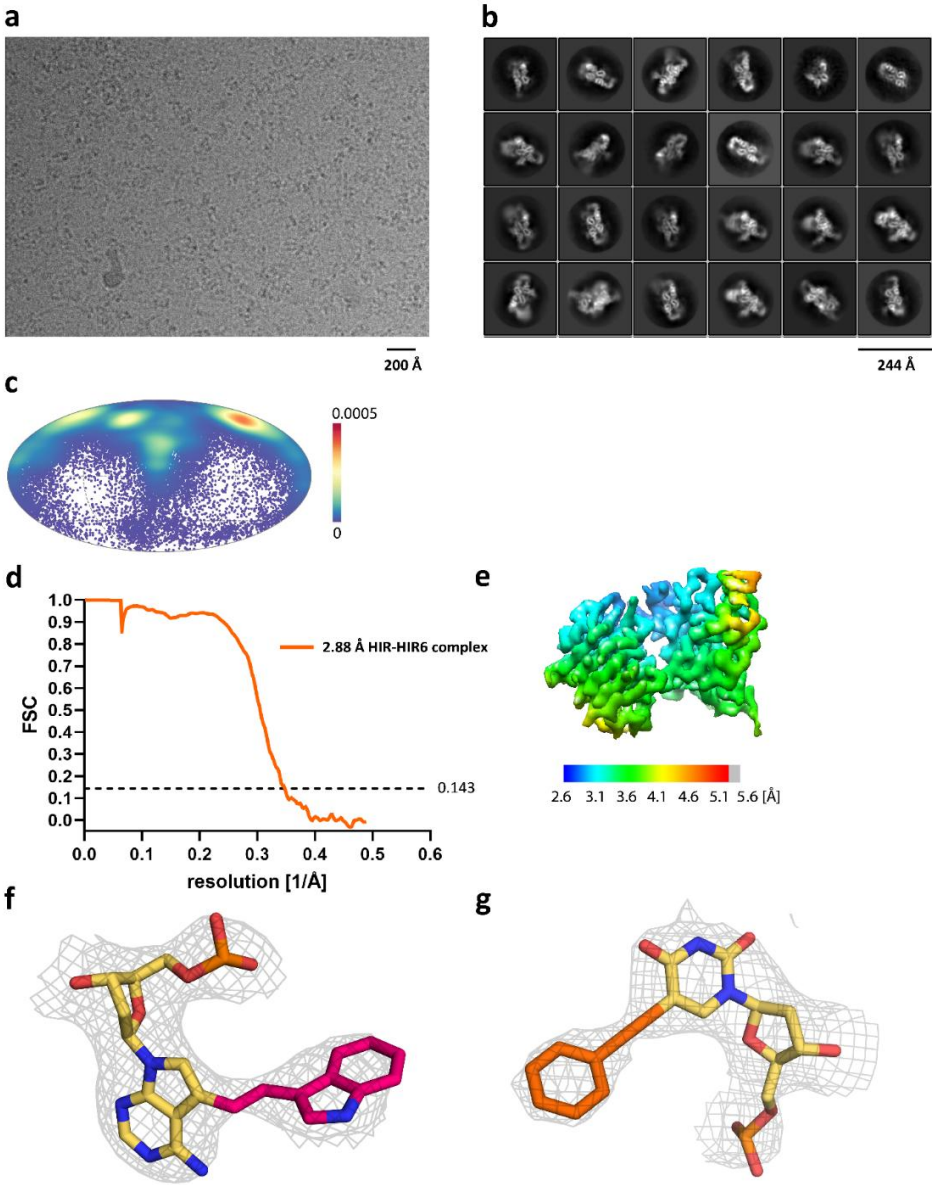


Figure S24. Cryo-EM of HIR-**HIR-6** complex. a) Micrograph of the complexes in free-standing ice after MotionCor2[31] correction at defocus of ~2.5 µm. b) 2D-class averages of HIR-**HIR-6** complexes. c) Angular distribution for particles of a on globe-like plane. d) Fourier shell correlation (FSC) curves for the HIR-**HIR-6** complex. The plot of the FSC between two independently refined half-maps shows the overall resolution of the two maps as indicated by the gold standard FSC 0.143 cut-off criterion.[32] e) Surface representation of local resolution distribution focused on the FnIII-1, L2' and **HIR-6** part of the HIR-**HIR-6** complex. The map is colored according to the local resolution calculated within the RELION software package. Resolution is as indicated in the color bar. f,g) Examples of the LocScale cryo-EM maps (isomesh level 55) depicting the modified DNA bases of the **HIR-6** aptamer.

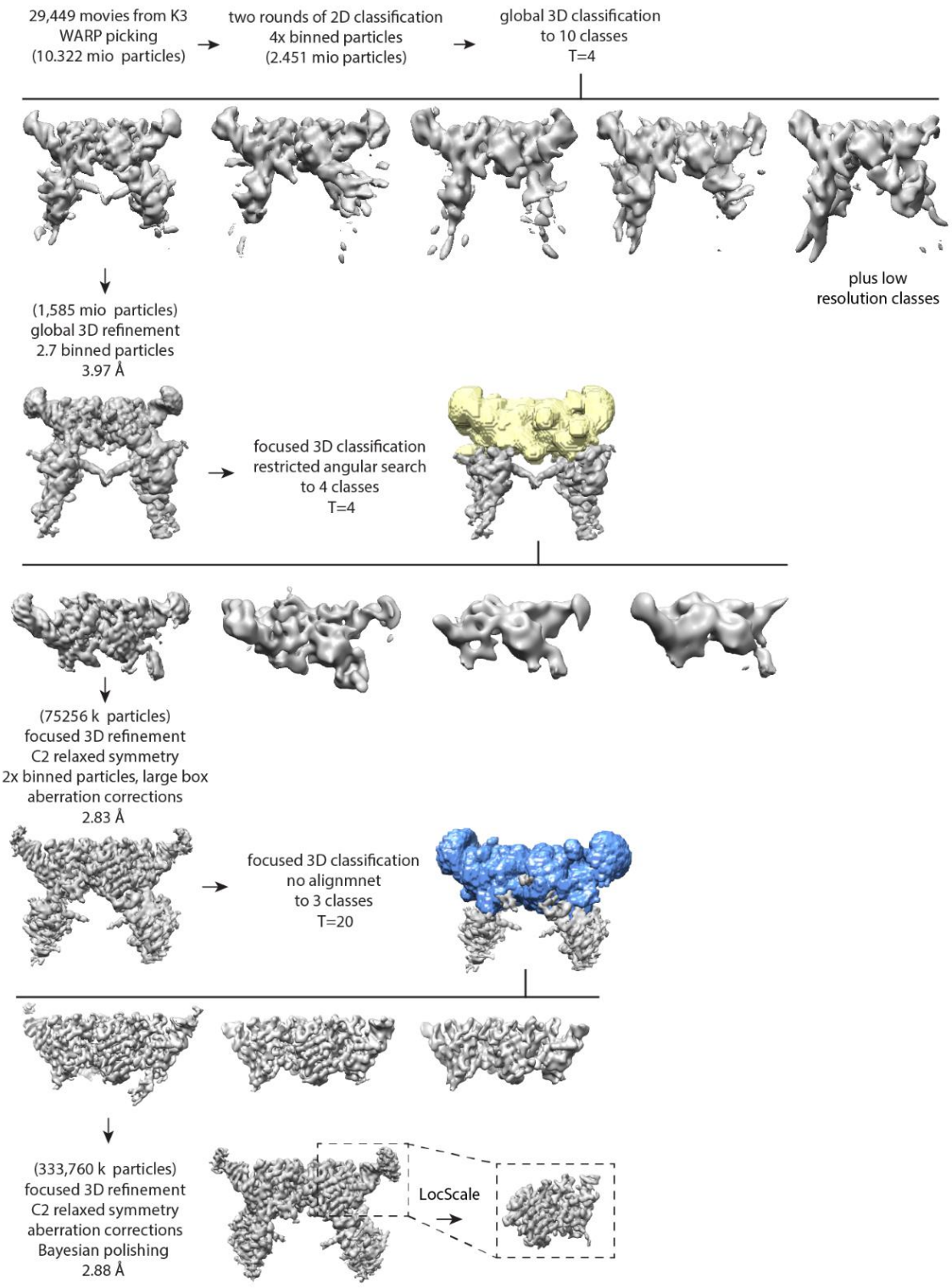


Figure S25: Cryo-EM data processing workflow for HIR-**HIR-6** complex. Cryo-EM single particle analysis data processing workflow of the HIR-**HIR-6** complex. The workflow shows a summary of the 3D classification and refinement scheme, together with representative focused classification masks.

Table S10: Cryo-EM structure determination and validation statistics.

Name of structure	**HIR-HIR6 complex**
PDB ID	**9SA8**
EMDB ID	**EMD-54689**
Data collection and processing	
Microscope	Titan Krios
Voltage (kV)	300
Camera	Gatan K3 BioQuantum
Magnification (x)	165,000
Nominal defocus range (negative µm)	0.5-3.0
Exposure time (s)	2.0
Electron exposure (e^{-}/Å^2)	51
Number of frames collected (no.)	40
Number of frames processed (no.)	25
Pixel size (Å)	0.5113
Micrographs (no.)	29,449
Total particle images (no.)	10,322,244
Refinement	
Particles per class (no.)	333,760
Map resolution (Å), 0.143 FSC	2.88
Map sharpening B factor (Å^2)	-19.9
Map versus model cross-correlation	0.89
Map versus model cross-correlation for ligands	0.90
Model composition	
Non-hydrogen atoms	2,704
Protein residues	274
Nucleotide residues	10
Chains	3
Ligands	9
B factors (Å^2) *min/max/mean*	
Protein	26.33/129.54/59.18
Nucleotide	42.28/168.77/106.46
Ligand	33.81/106.82/57.50
R.m.s. deviations	
Bond lengths (Å)	0.005
Bond angles (°)	1.029
Validation	
MolProbity score	1.92
All-atom clashscore	5.96
Rotamer outliers (%)	3.27
Ramachandran plot	
Favored (%)	96.62
Allowed (%)	3.38
Outliers (%)	0

10.4 Cryo-EM model building and refinement

The model is limited to the HIR domain FnIII-1 and L2', where the **HIR-6** aptamer interacts with the protein moiety. Specifically, modelled residues span FnIII-1/469-519, FnIII-1/527-573, FnIII-1/578-592 and L2'/308-468. The model for the **HIR-6** aptamer is limited to residues 20 to 38, the cryo-EM density map for the rest of the aptamer is not interpretable, only showing low resolution secondary structure features.

The X-ray structure of the HIR ectodomain (PDB ID 4zbx)[27] was used as a starting model. One protomer of the HIR ectodomain was docked into the cryo-EM map by Molrep[33] and only domain L2' and FnIII-1 were retained for further modeling. The model for **HIR-6** aptamer was modeled manually in Coot 0.8.9.1[34] together with adjustments to the HIR ectodomain protein moiety. Refinement in Coot was done against a blurred MTZ map (blurring B = 50 Å^2) generated in CCP-EM[28]. The aptamer register of residues 20 to 38 was verified by cryo-EM density matching. Ligand geometry restraints for modified DNA bases were generated using Acedrg[35] within the CCP4i2 package[36]. Atom naming of the modified nucleotides was matched with the respective unmodified nucleotides from the monomer library, and the “NON-POLYMER” identifier was manually changed to “DNA” to ensure that the monomer units seamlessly link into the DNA chain. Model self-restraints were used, as well as base pairing and parallelity restraints for DNA, which were automatically generated by the program libG[37] running under Refmac 5.8.0405[38] within the CCP4 Interface 8.0.010[39] and curated manually. The final model was refined in real space[40] against a post-processed MRC map in Phenix 1.21-5207[41], using self-restraints with the strict rotamer matching option enabled, as well as secondary structure restraints, including base pairing and parallelity restraints for DNA. The restraints were generated automatically in Phenix 1.21-5207[41]. The final refinement round in Phenix included one cycle of ADP refinement only. The refined models were validated using MolProbity[42] and the wwPDB database[43] validation server.

Structure analysis was performed using ChimeraX[44], Coot[45], and PISA[46] server. Non-covalent interactions were identified in ChimeraX with relaxed geometry criteria enabled. ChimeraX and Pymol[47] were used for structure visualization.

10.5 HIR-6 interactions

Table S11. HIR-6 aptamer intra-structure interactions.

Canonical base pairing				
C20	base	G38	base	Watson-Crick base-pair hydrogen bonds
G21	base	C37	base	
C22	base	G36	base	
G23	base	C34	base	
A33	base	U25	base	
A32	base	U26	base	
Canonical base π-π stacking				
C20	base	G21	base	π-π stacking
G21	base	C22	base	
C22	base	G23	base	
G23	base	G36	base	
U25	base	U26	base	
A27	base	A28	base	
A28	base	C29	base	
A32	base	A33	base	
G36	base	C37	base	
C37	base	G38	base	
Non-canonical hydrogen bonds				
A27	base	A35	base	Hydrogen bond
A33	base	A35	base	
U25	base	A27	indole	
C29	base	A32	deoxyribose O4′	
Other π interactions				
G23	base	A35	indole	π-π stacking
U25	base	A35	indole	
A33	base	A35	indole	
G30	base	U31	phenyl	
A32	base	A33	indole	T-shape π-π stacking
C29	base	U31	deoxyribose O4′	Lone electron pair - π interaction
C34	deoxyribose O4′	A35	base	
U26	deoxyribose O4′	A27	indole	
U26	deoxyribose C5′	A27	indole	CH - π interaction
U25	deoxyribose C5′	U26	phenyl	

Table S12. HIR-6-HIR interactions.

Hydrogen bonds				
U25	base	R488	side chain	Hydrogen bond
A33	base	N547	side chain	
G30	base	T320 (L2′)	side chain	
U31	base	R554	side chain	
G30	base	T320 (L2′)	main chain	
U31	base	L552	main chain	
Salt bridges				
C22	phosphate	K544	side chain	Salt bridge
U26	phosphate	R479	side chain	
U26	phosphate	R479	side chain	
A28	phosphate	K484	side chain	
π interactions				
U31	base	W551	side chain	π-π stacking
U26	phenyl	Y477	side chain	
U25	phenyl	H548	side chain	T-shape π-π stacking
G30	base	E318-K319 (L2′)	main chain	Peptide bond π-π stacking
A32	indole	G550-W551	main chain	
U24	phenyl	Q546-N547	main chain	
U26	base	L486	side chain	CH - π interaction
G30	base	E318 B	side chain	
A28	indole	R479	side chain	
A28	indole	K484	side chain	
A33	indole	L538	side chain	
U31	phenyl	R554	side chain	Cation – π interaction

11. Copies of NMR spectra of prepared compounds

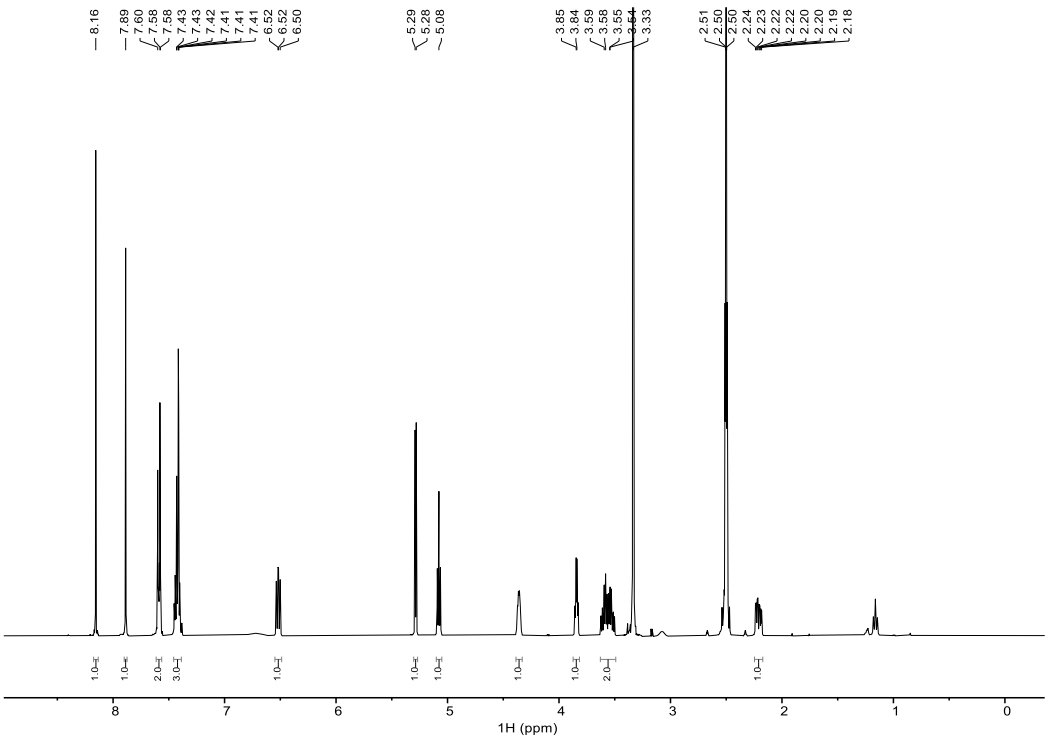


Figure S26. ^{1}H NMR spectrum of **dA^{EPh}**.

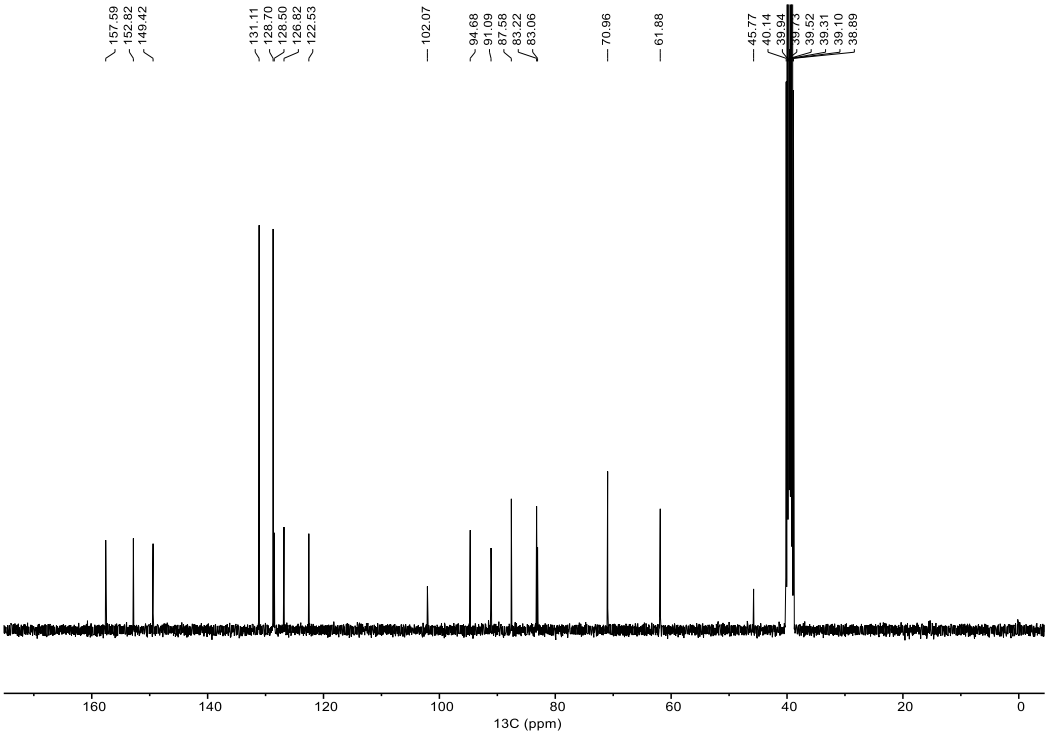


Figure S27. ^{13}C NMR spectrum of **dA^{EPh}**.

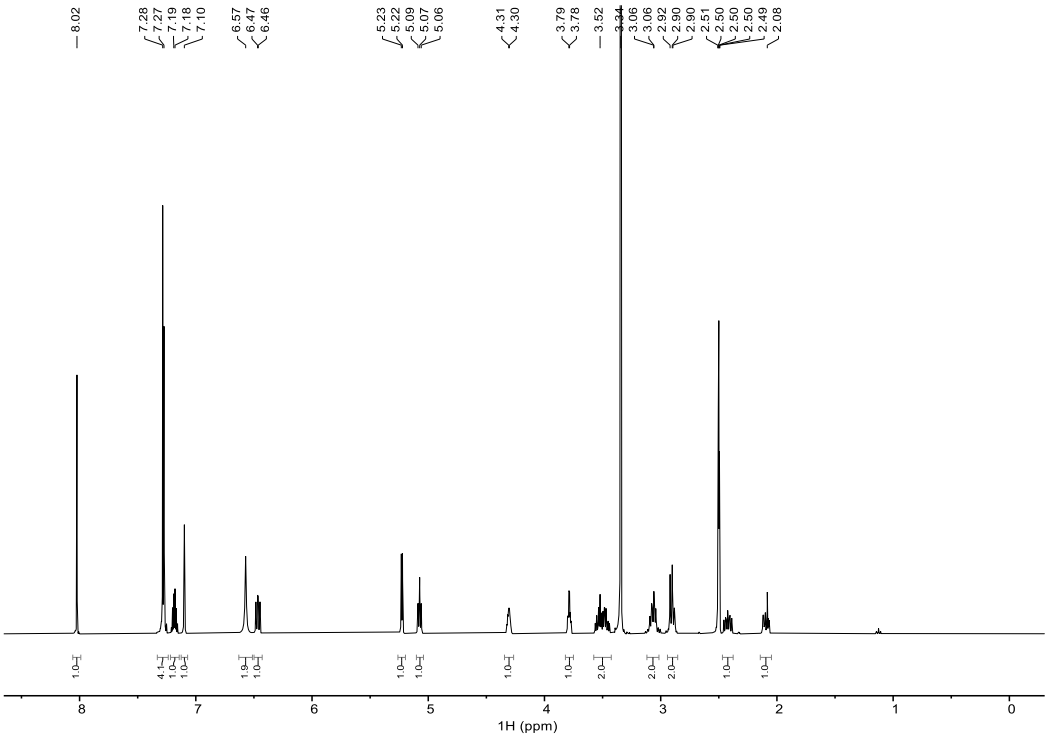


Figure S28. ^{1}H NMR spectrum of **dA^{APh}**.

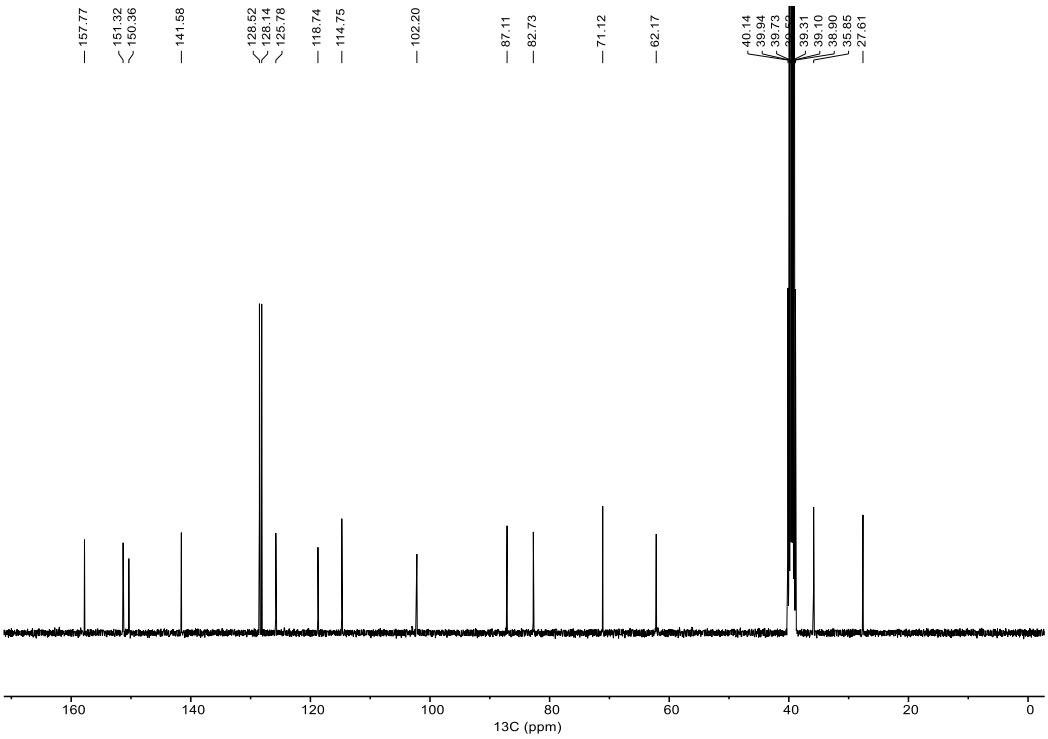


Figure S29. ^{13}C NMR spectrum of **dA^{APh}**.

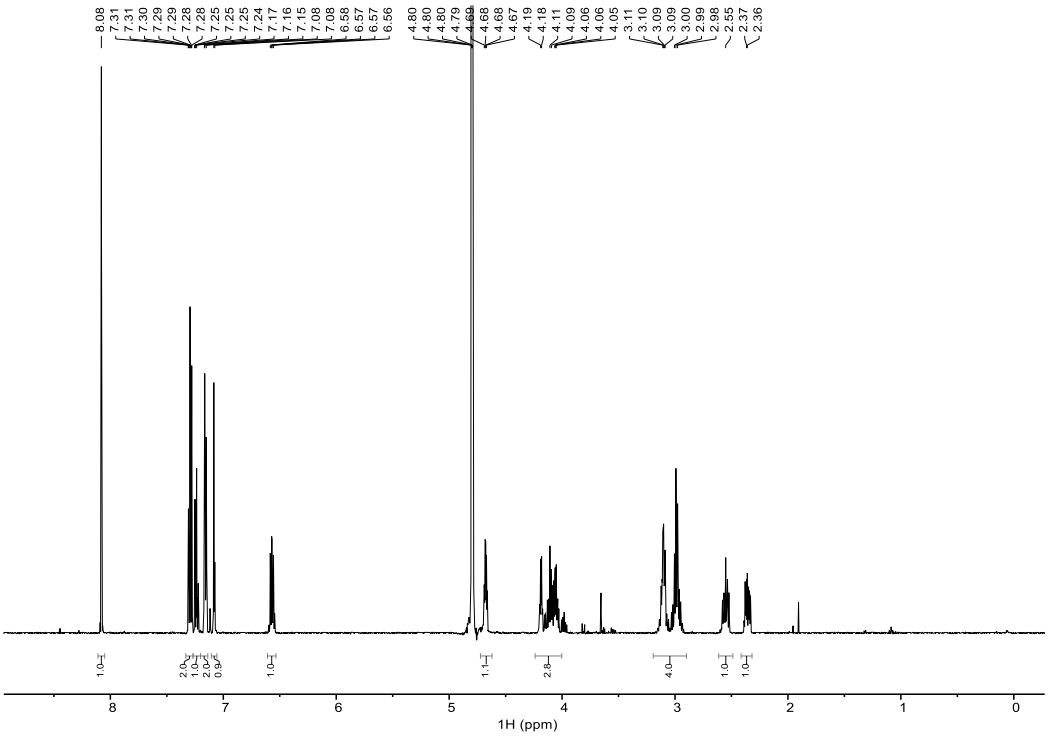


Figure S30. ^{1}H NMR spectrum of **dA^{APh}TP**.

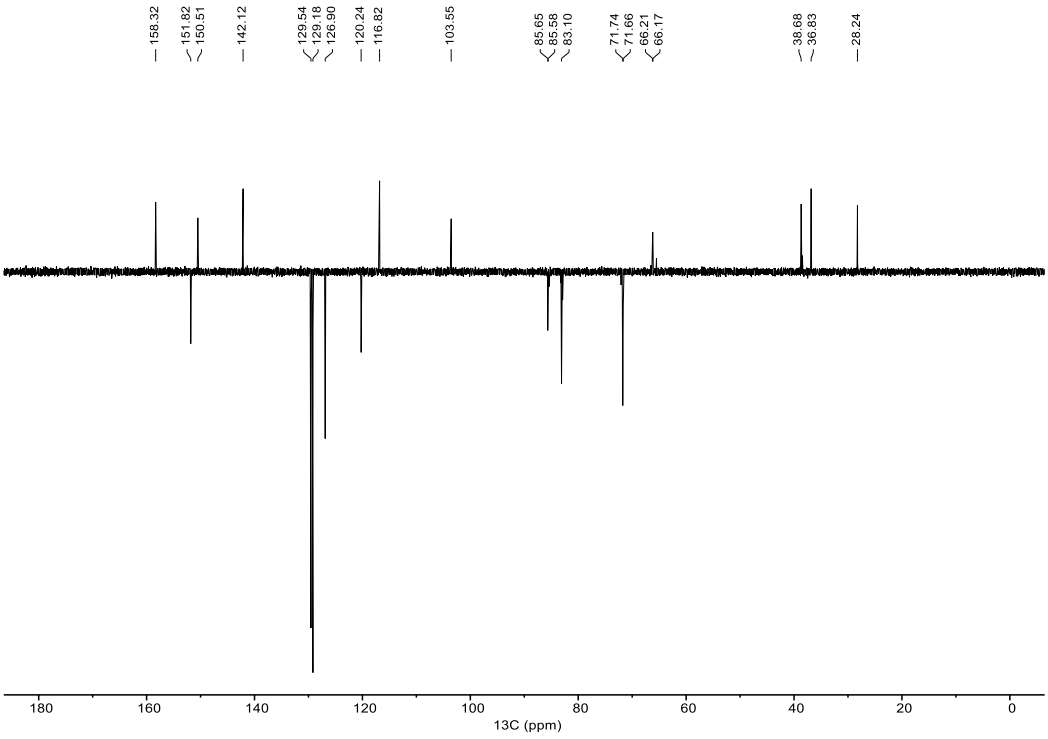


Figure S31. ^{13}C APT NMR spectrum of **dA^{APh}TP**.

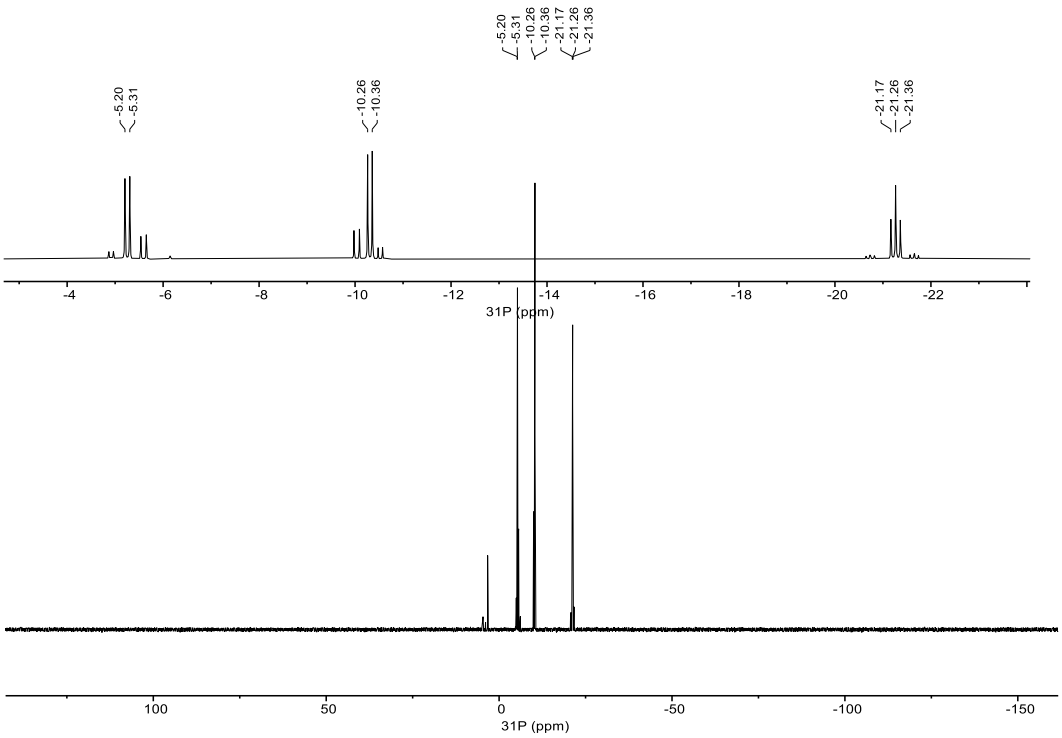


Figure S32. ^{31}P NMR spectrum of **dA^{APh}TP**.

12. MS characterization of synthesized modified ONs

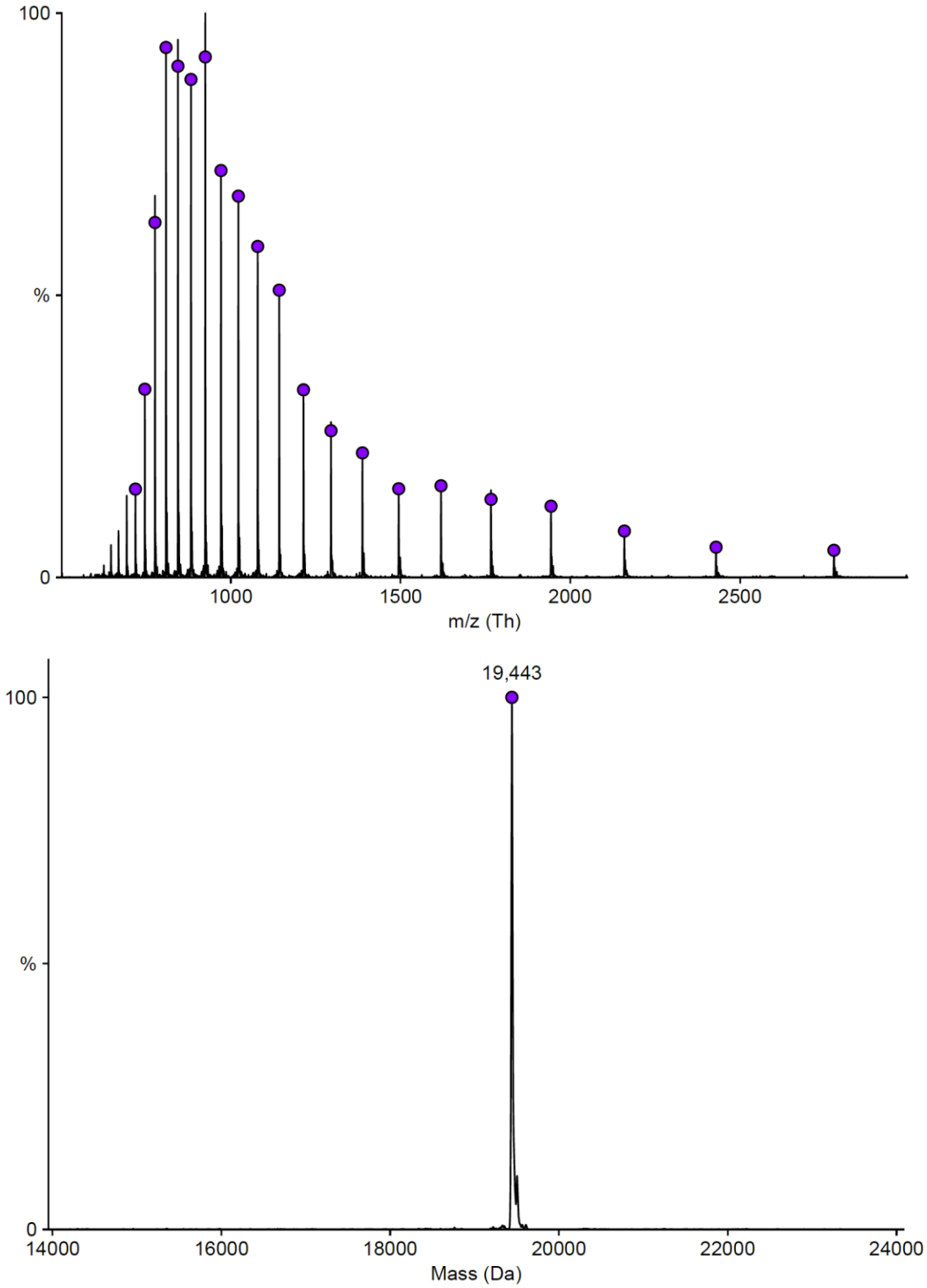


Figure S33. Raw and deconvoluted MS spectrum of 5´-Cy5-**HIR-6**, calculated mass: 19445 Da, found mass: 19443 Da.

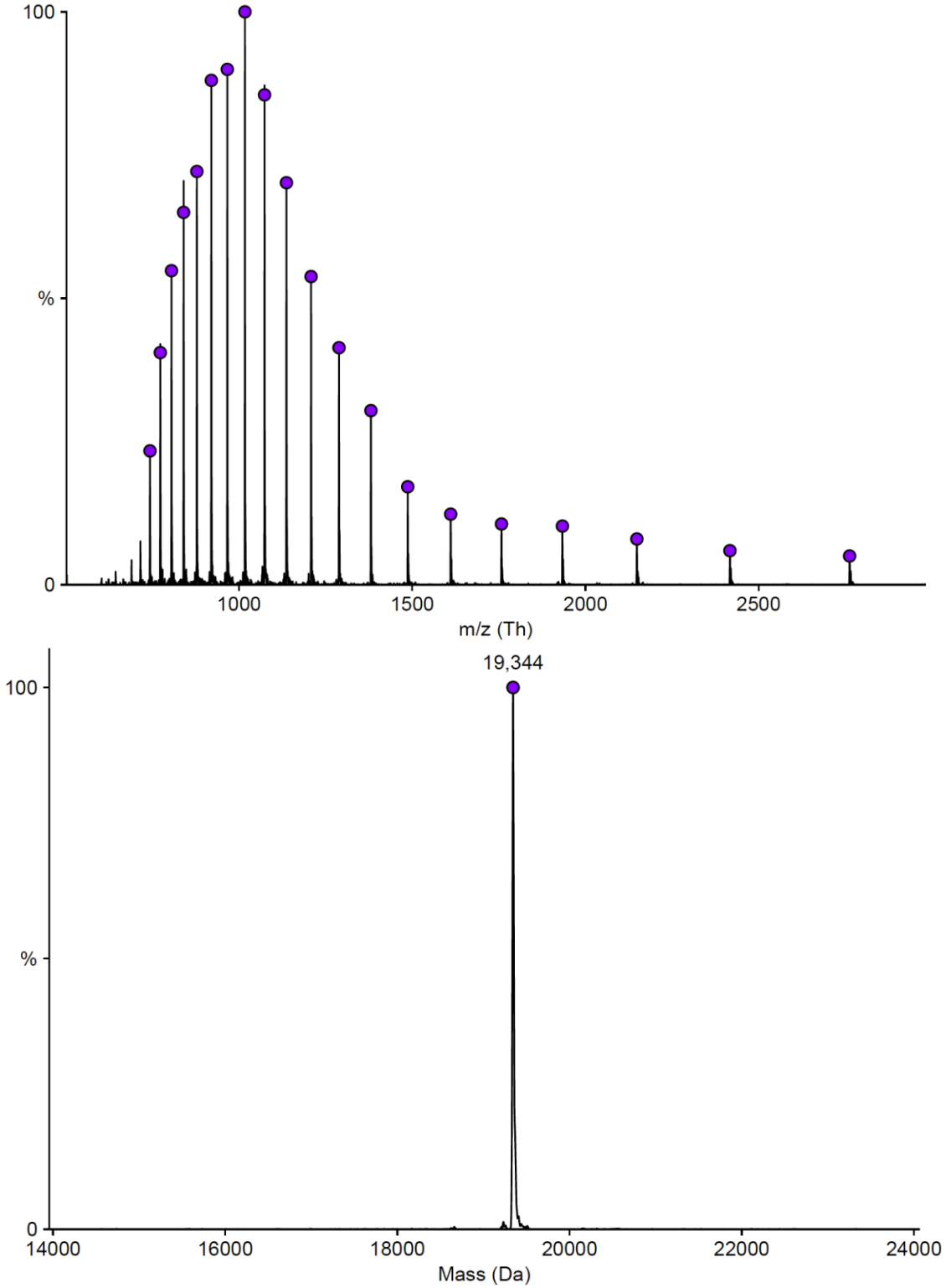


Figure S34. Raw and deconvoluted MS spectrum of 5´-Biotin-**HIR-6**, calculated mass: 19349 Da, found mass: 19344 Da.

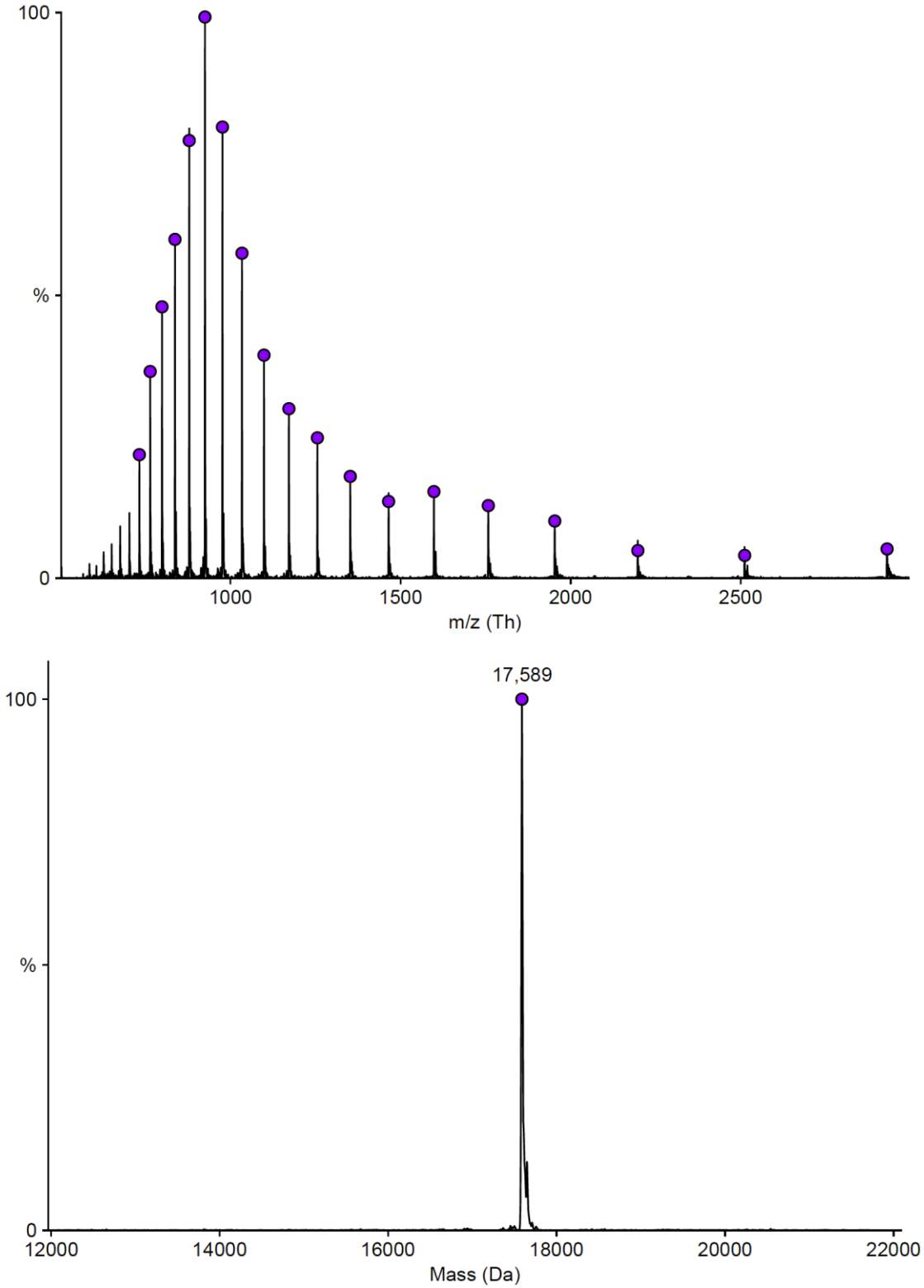


Figure S35. Raw and deconvoluted MS spectrum of 5´-Cy5-**HIR-6_T1**, calculated mass: 17591 Da, found mass: 17589 Da.

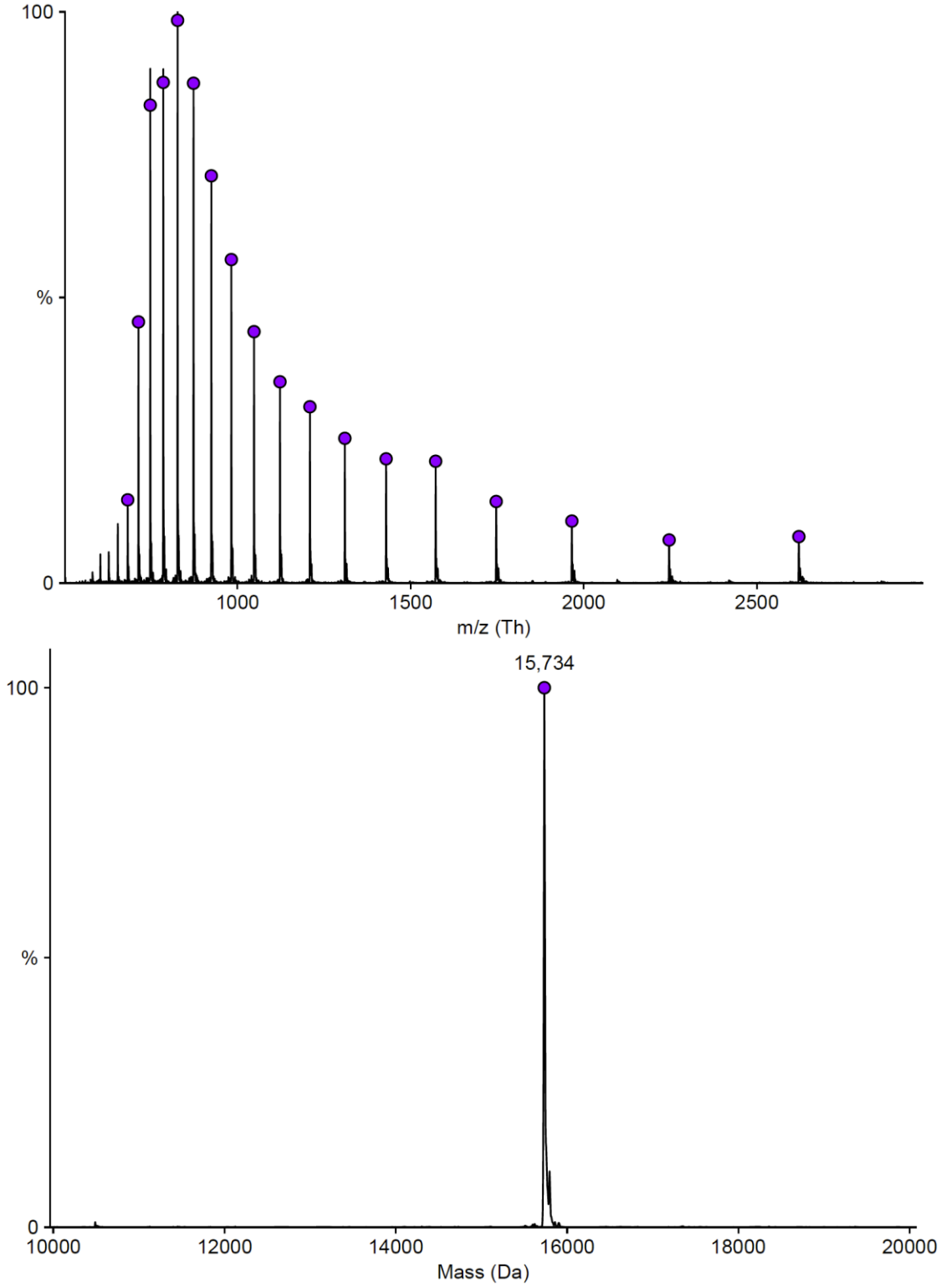


Figure S36. Raw and deconvoluted MS spectrum of 5´-Cy5-**HIR-6_T2**, calculated mass: 15737 Da, found mass: 15734 Da.

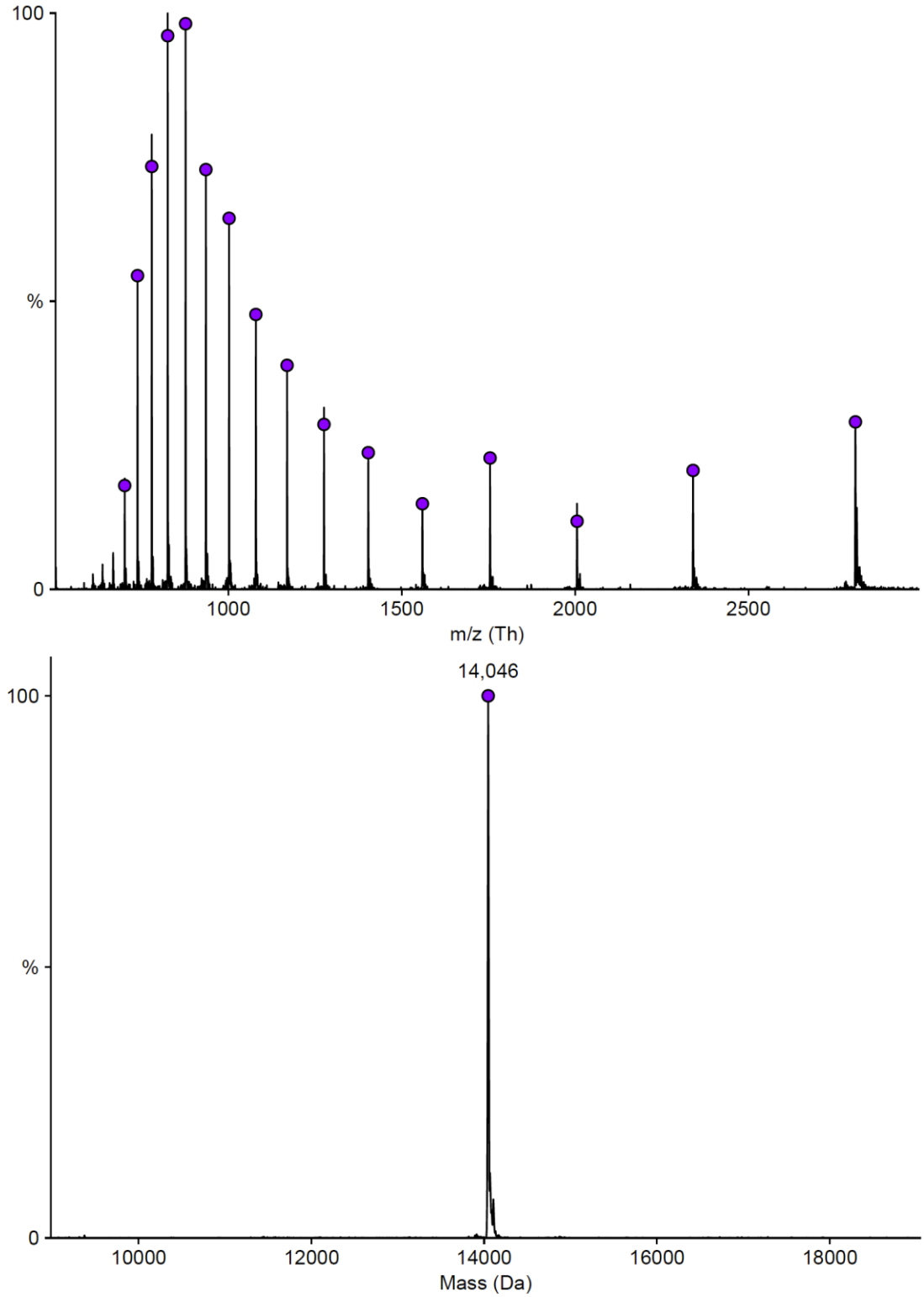


Figure S37. Raw and deconvoluted MS spectrum of 5´-Cy5-**HIR-6_T3**, calculated mass: 14048 Da, found mass: 14046 Da.

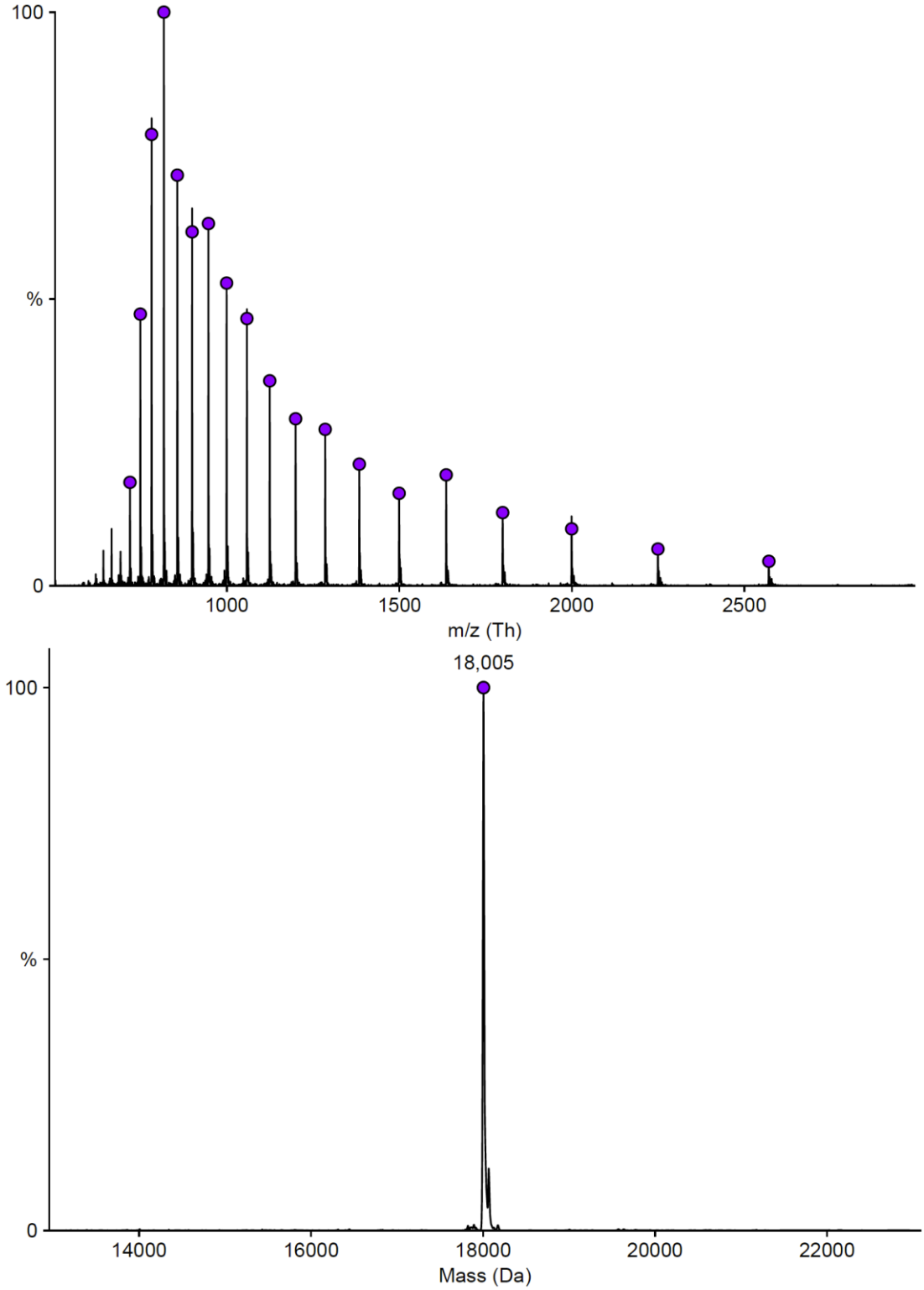


Figure S38. Raw and deconvoluted MS spectrum of 5´-Cy5-**HIR-6_T4**, calculated mass: 18008 Da, found mass: 18005 Da.

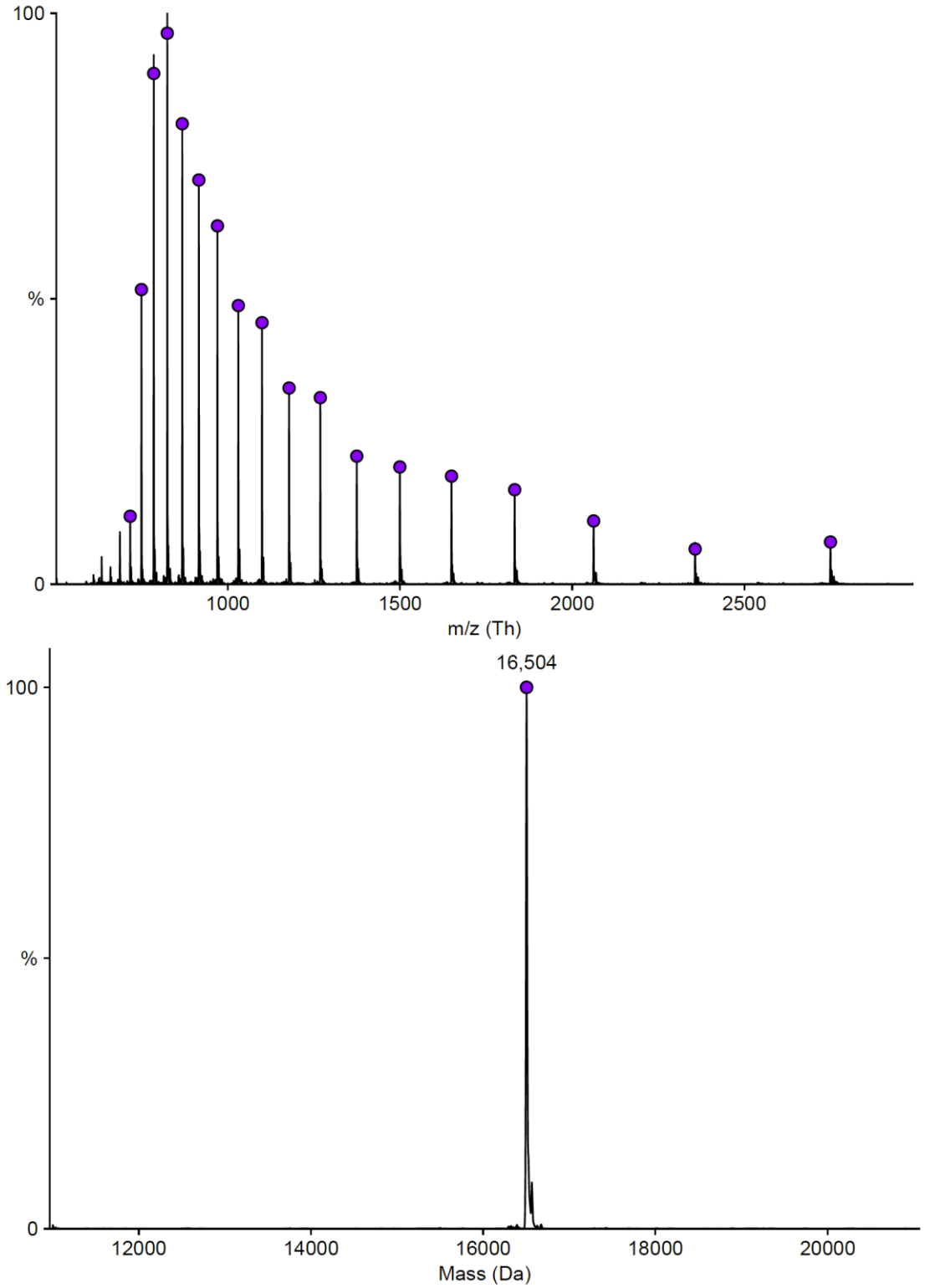


Figure S39. Raw and deconvoluted MS spectrum of 5´-Cy5-**HIR-6_T5**, calculated mass: 16507 Da, found mass: 16504 Da.

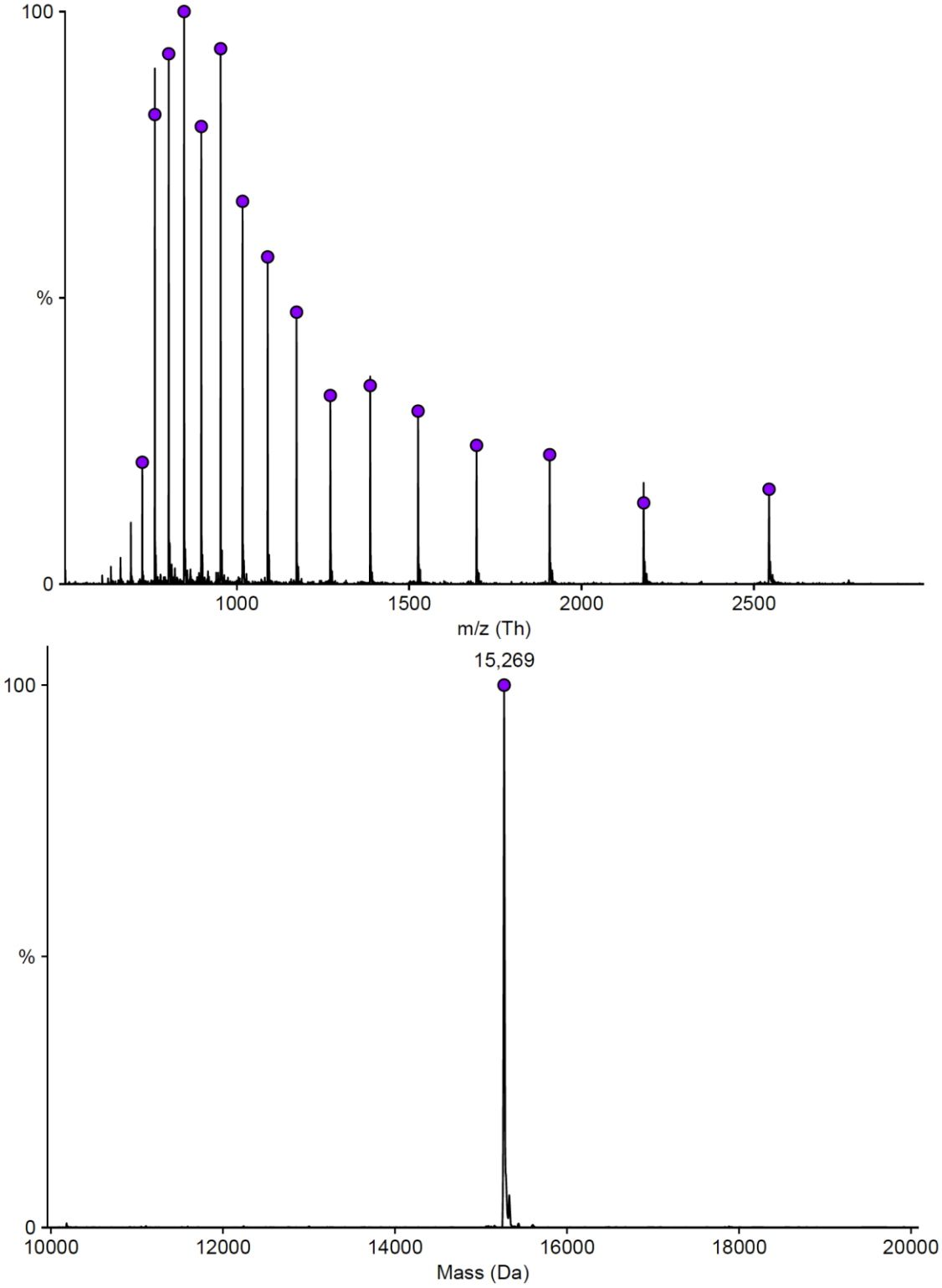


Figure S40. Raw and deconvoluted MS spectrum of 5´-Cy5-**HIR-6_T6**, calculated mass: 15271 Da, found mass: 15269 Da.

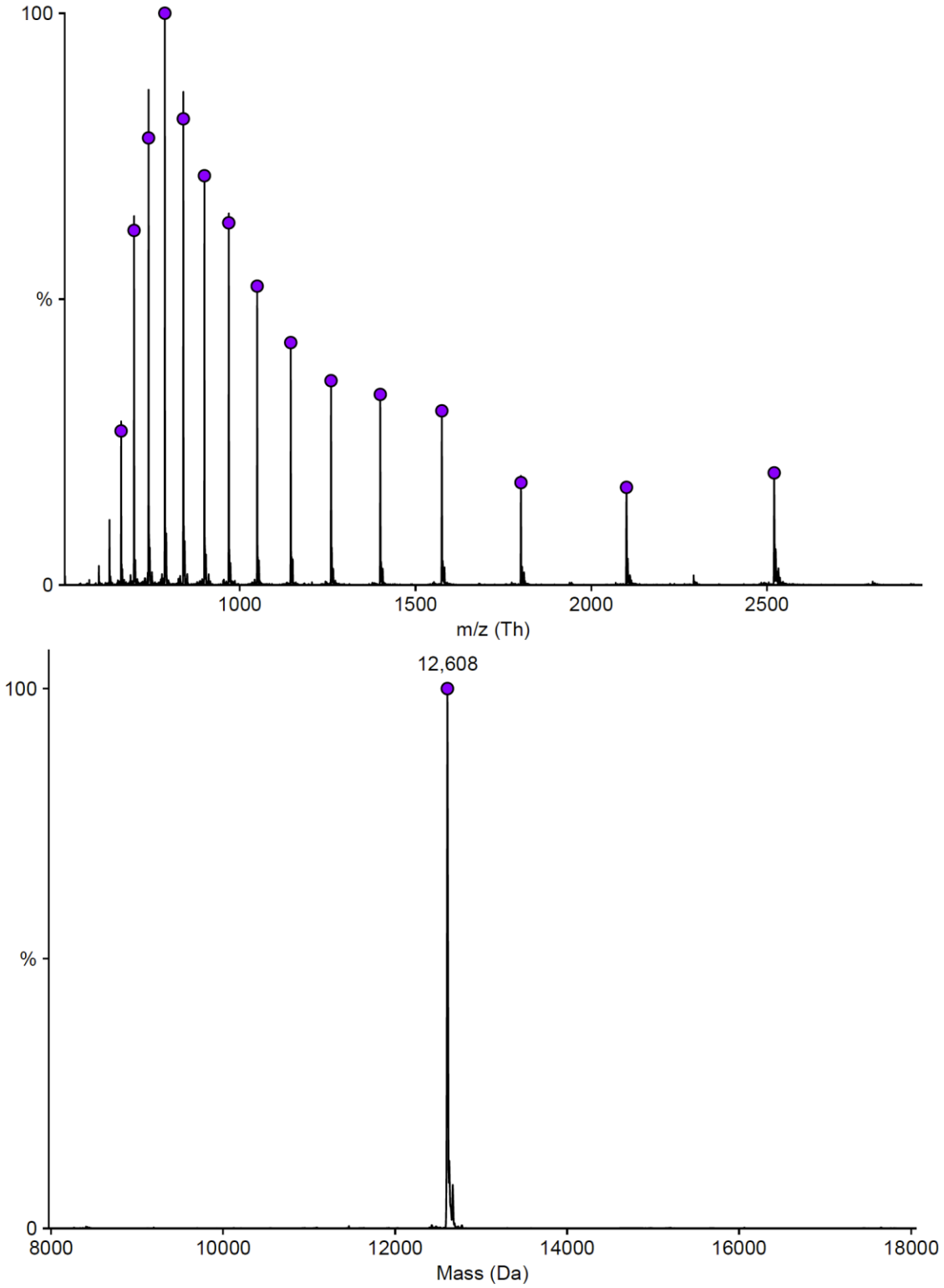


Figure S41. Raw and deconvoluted MS spectrum of 5´-Cy5-**HIR-6_T7**, calculated mass: 12611 Da, found mass: 12608 Da.

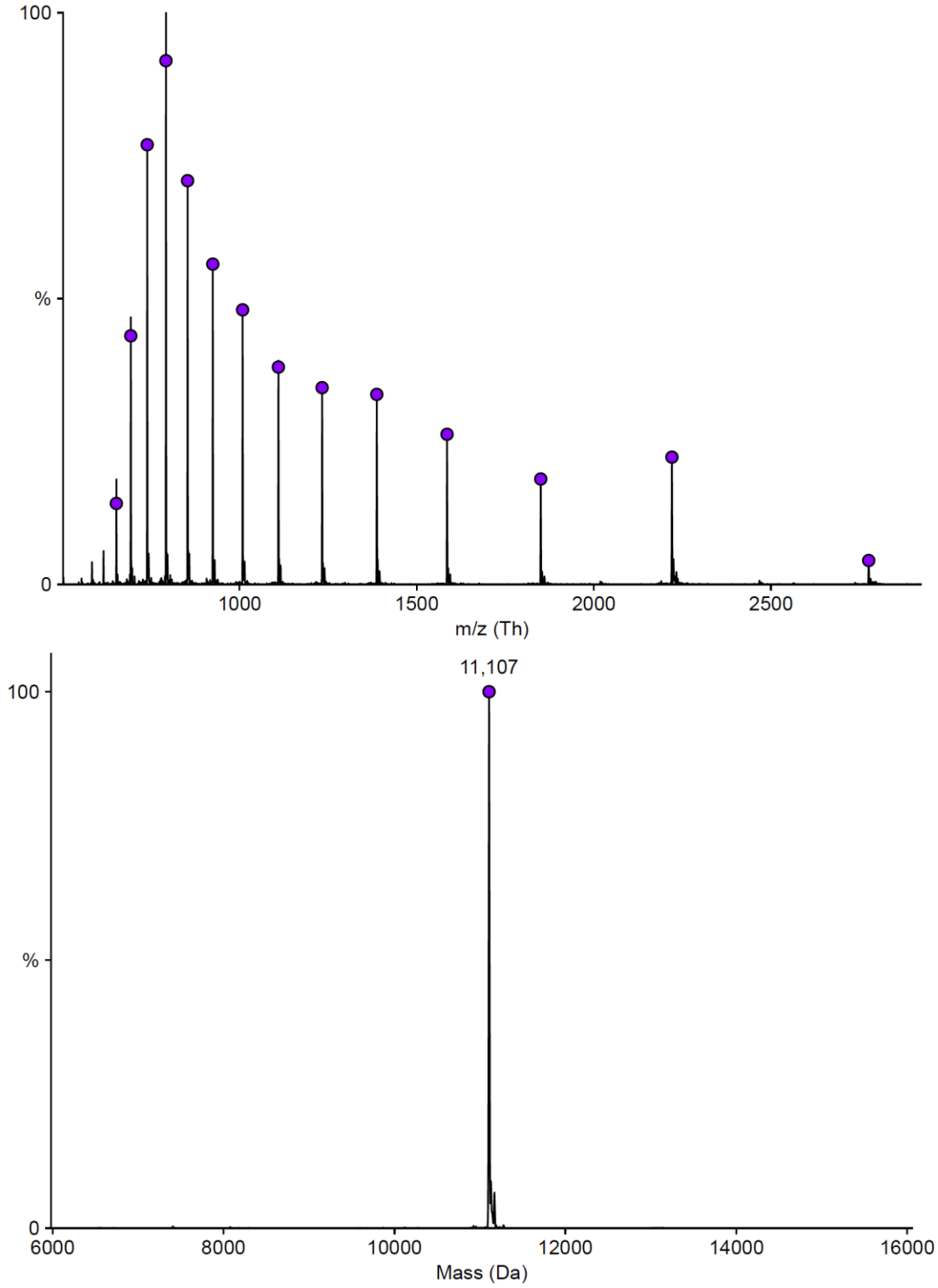


Figure S42. Raw and deconvoluted MS spectrum of 5´-Cy5-**HIR-6_T8**, calculated mass: 11110 Da, found mass: 11107 Da.

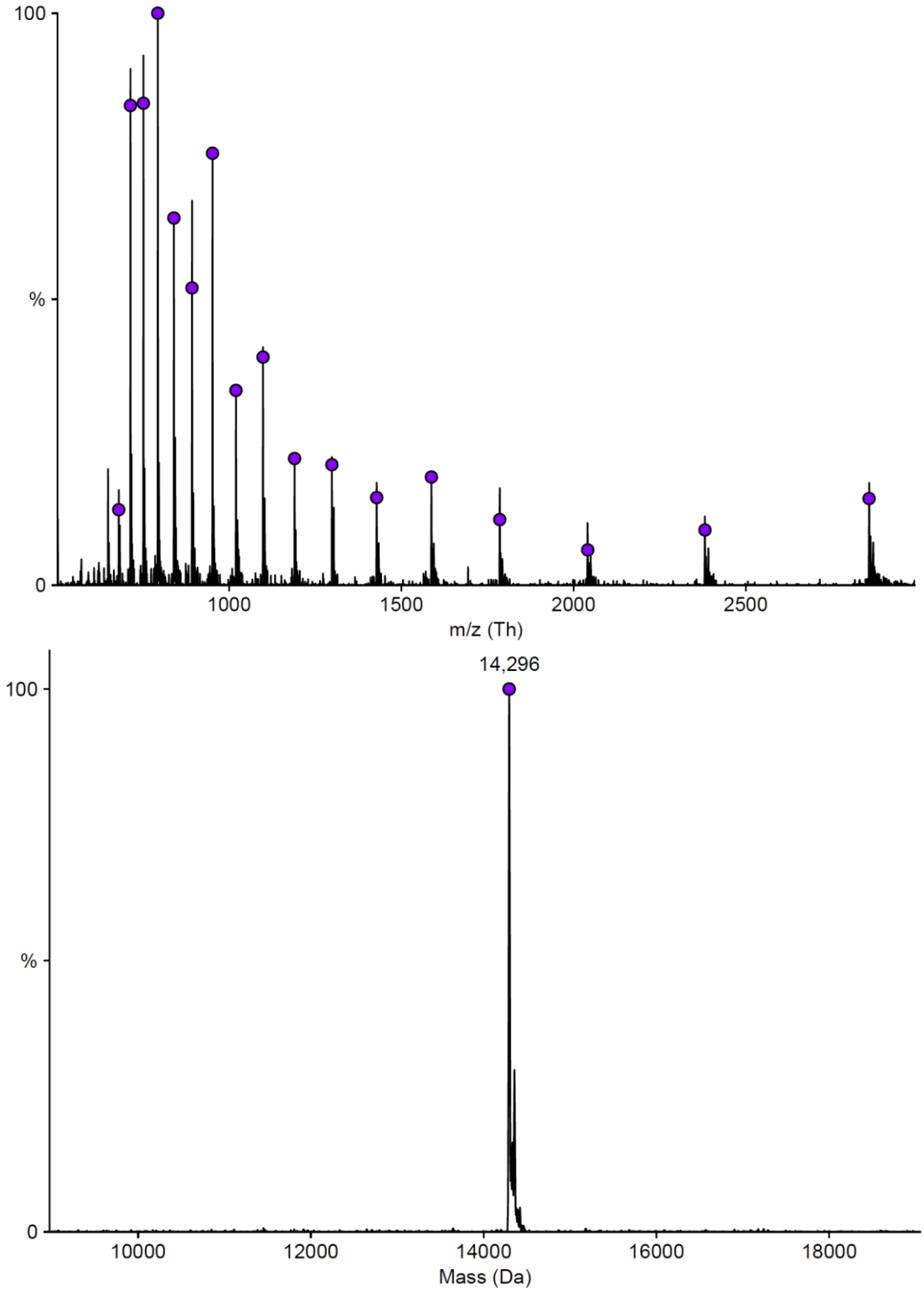


Figure S43. Raw and deconvoluted MS spectrum of 5´-Cy5-**HIR-6_T9**, calculated mass: 14299 Da, found mass: 14296 Da.

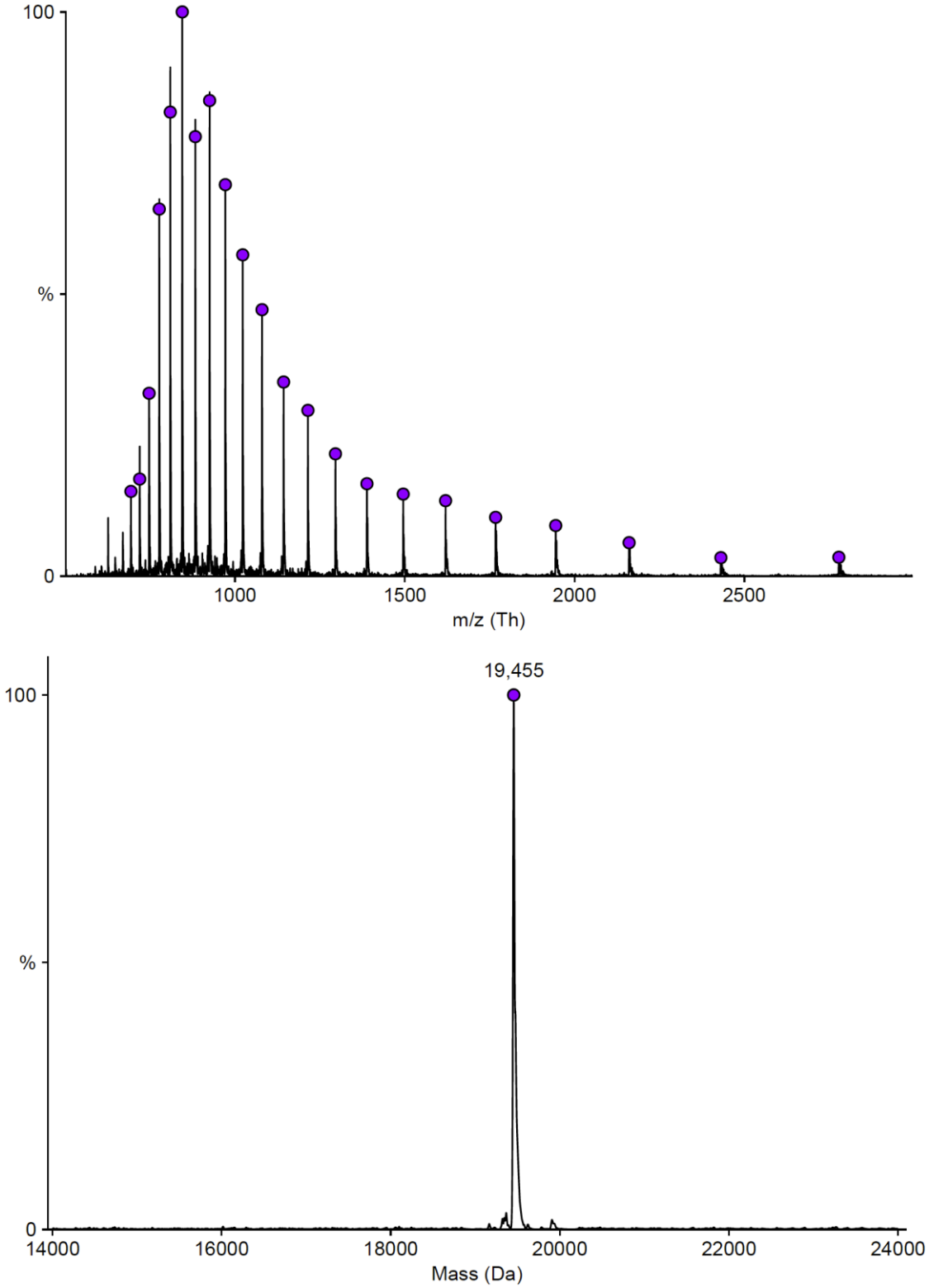


Figure S44. Raw and deconvoluted MS spectrum of 5´-Cy5-**HIR-6_V1**, calculated mass: 19457 Da, found mass: 19455 Da.

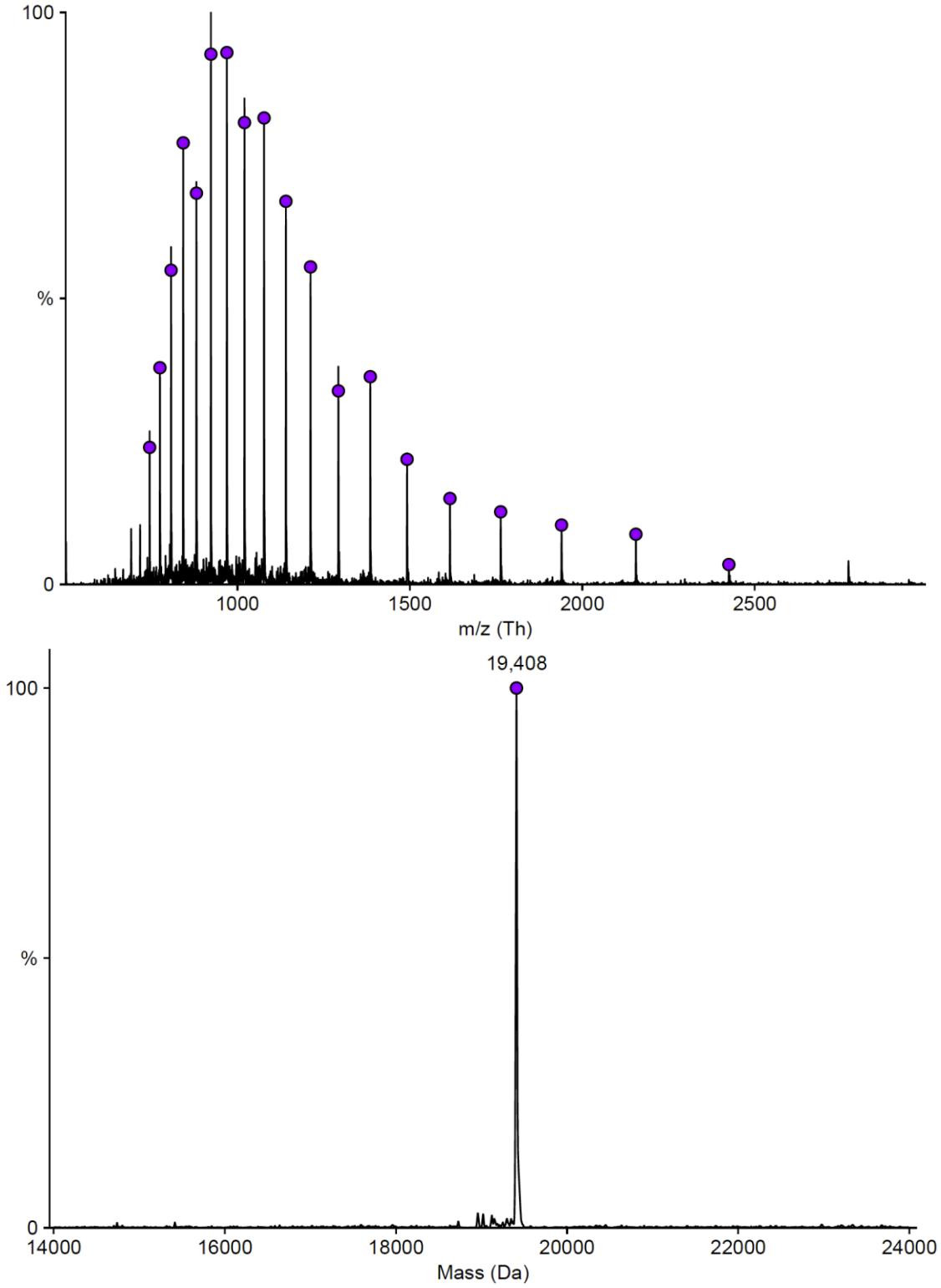


Figure S45. Raw and deconvoluted MS spectrum of 5´-Cy5-**HIR-6_V2**, calculated mass: 19413 Da, found mass: 19408 Da.

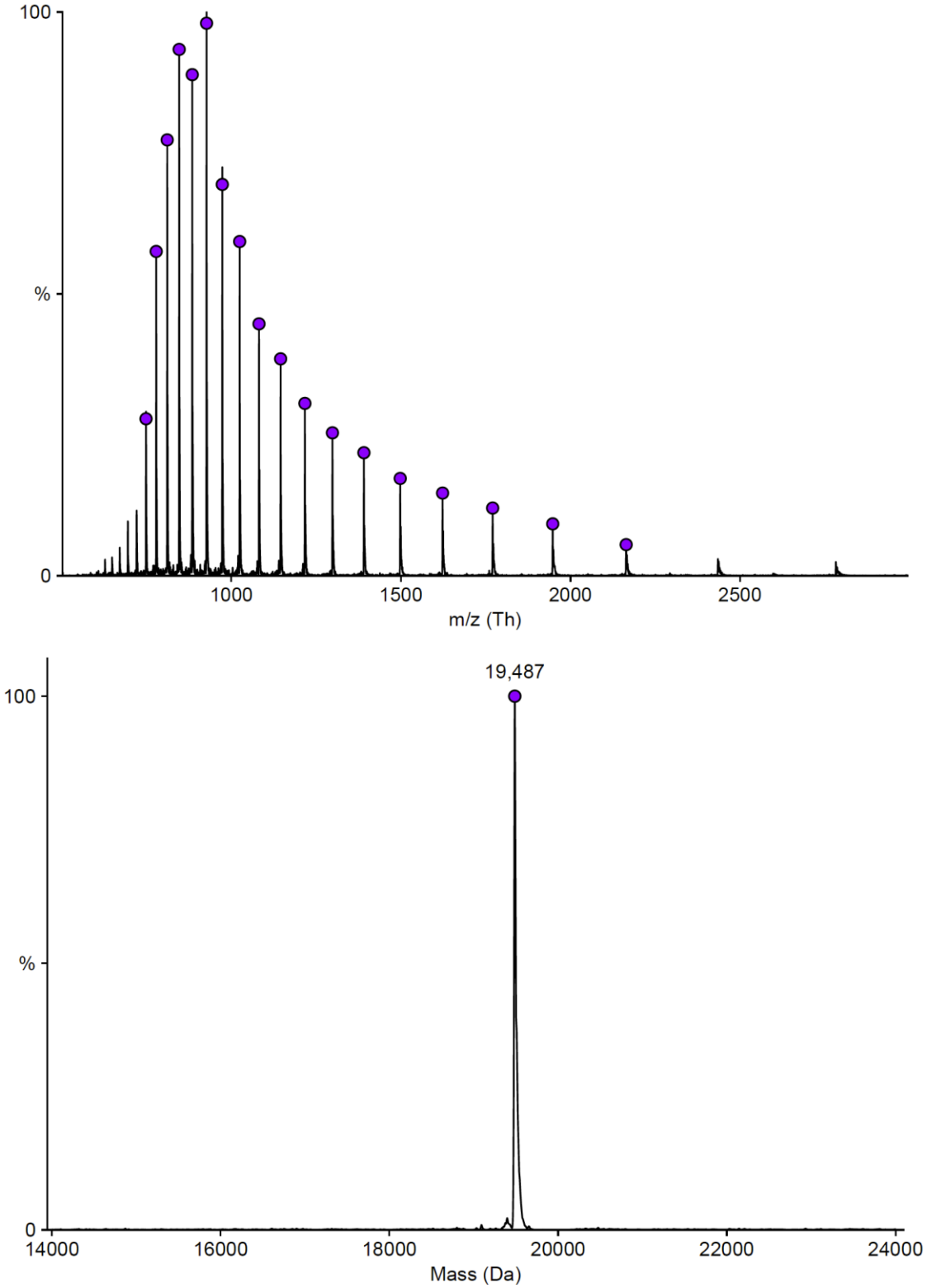


Figure S46. Raw and deconvoluted MS spectrum of 5´-Cy5-**HIR-6_V3**, calculated mass: 19490 Da, found mass: 19487 Da.

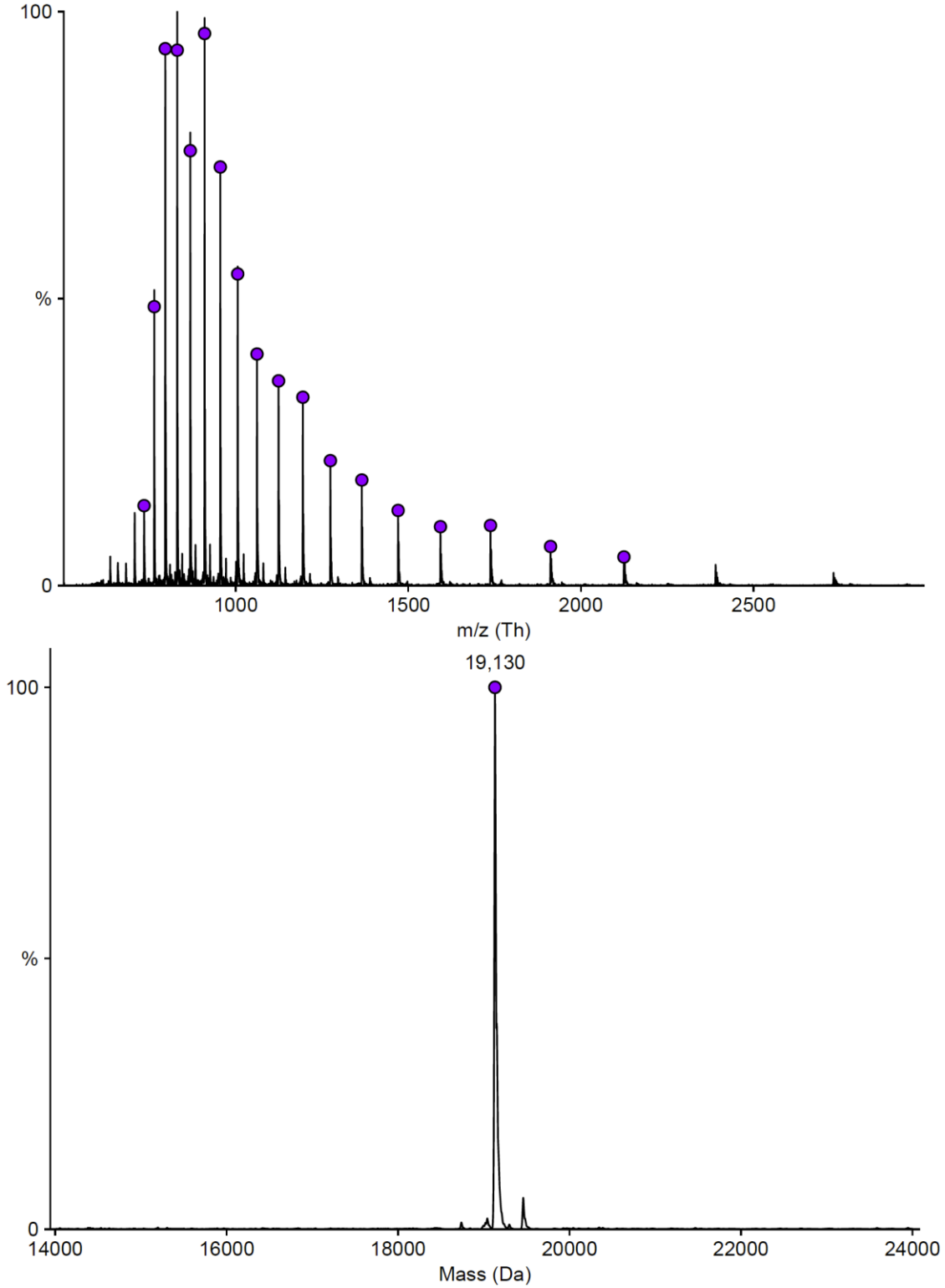


Figure S47. Raw and deconvoluted MS spectrum of 5´-Cy5-**HIR-6_V4**, calculated mass: 19133 Da, found mass: 19130 Da.

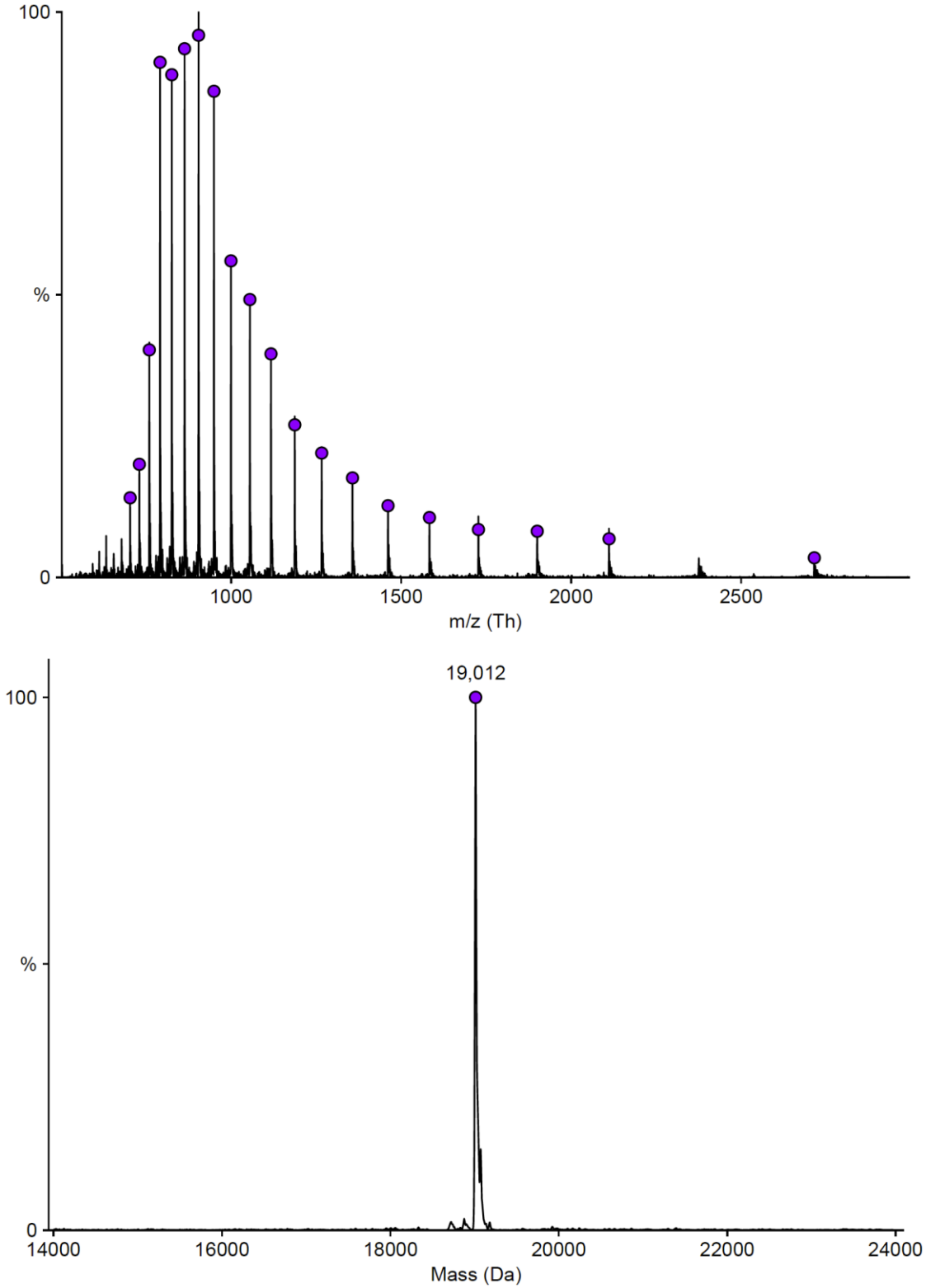


Figure S48. Raw and deconvoluted MS spectrum of 5´-Cy5-**HIR-6_M1**, calculated mass: 19014 Da, found mass: 19012 Da.

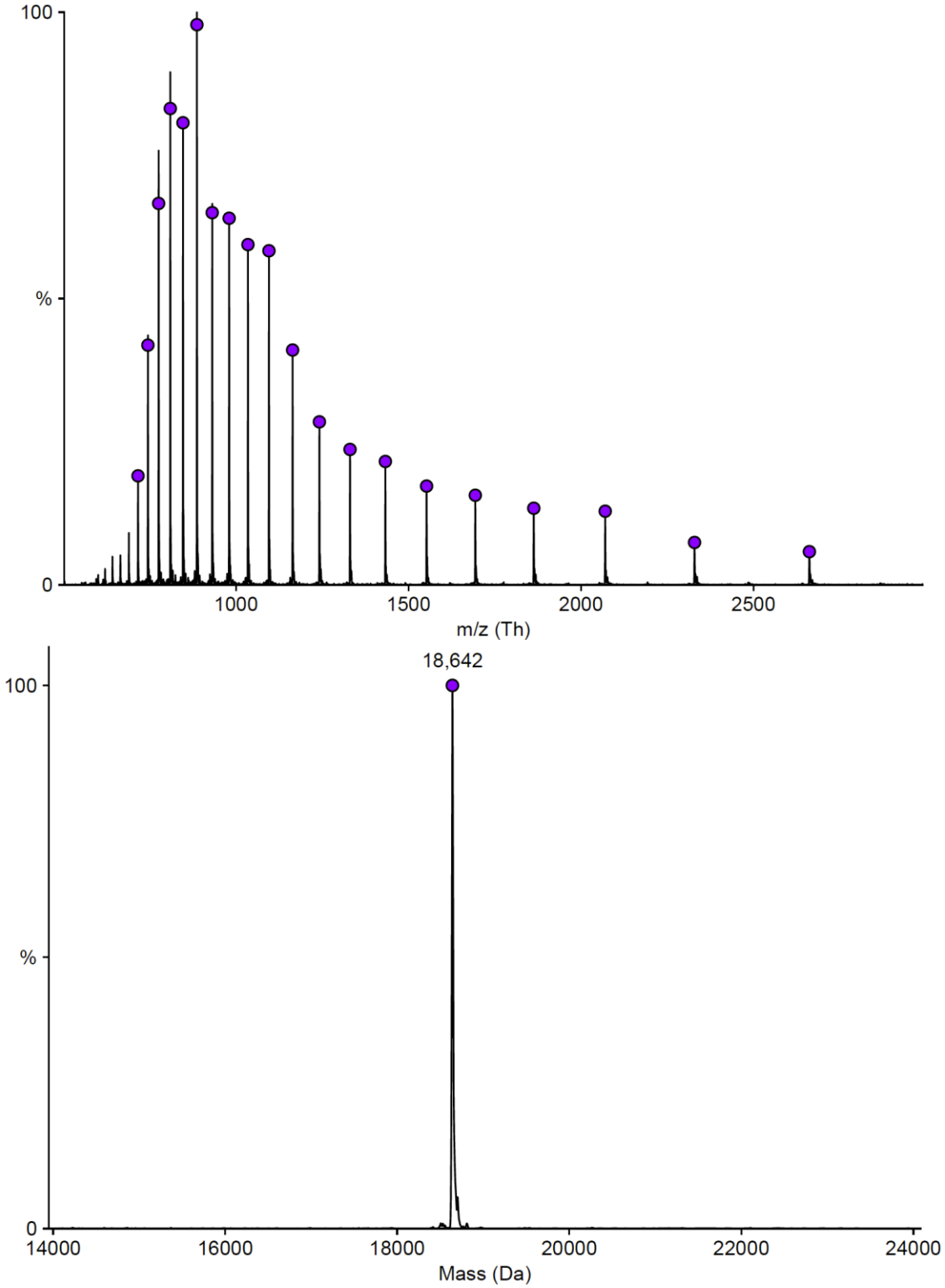


Figure S49. Raw and deconvoluted MS spectrum of 5´-Cy5-**HIR-6_M2**, calculated mass: 18644 Da, found mass: 18642 Da.

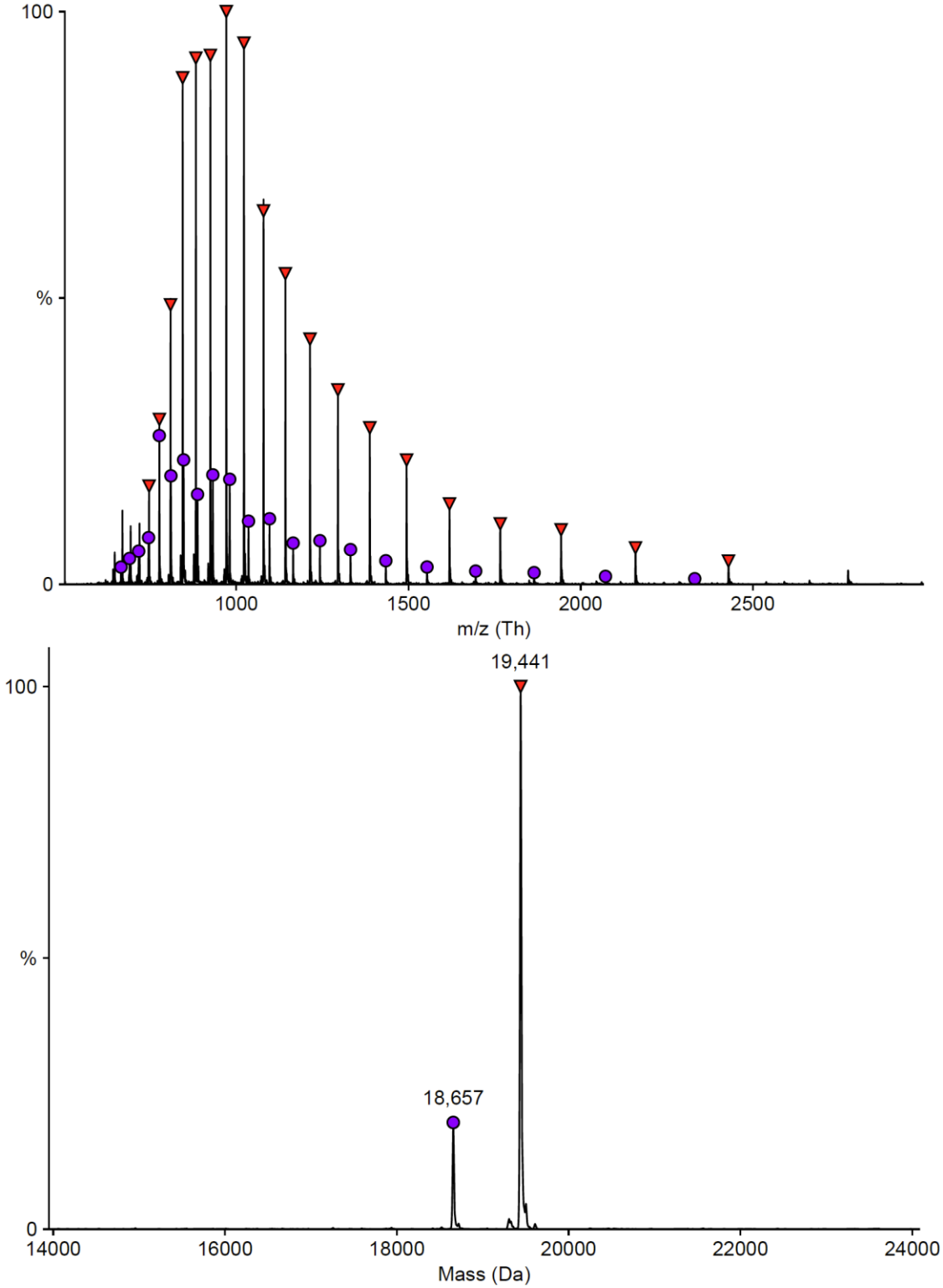


Figure S50. Raw and deconvoluted MS spectrum of 5´-Cy5-**HIR_SC**, calculated mass: 19445 Da, found mass: 19441 Da. The mass of 18657 Da corresponds to the N-2 product (calculated mass: 18661 Da).

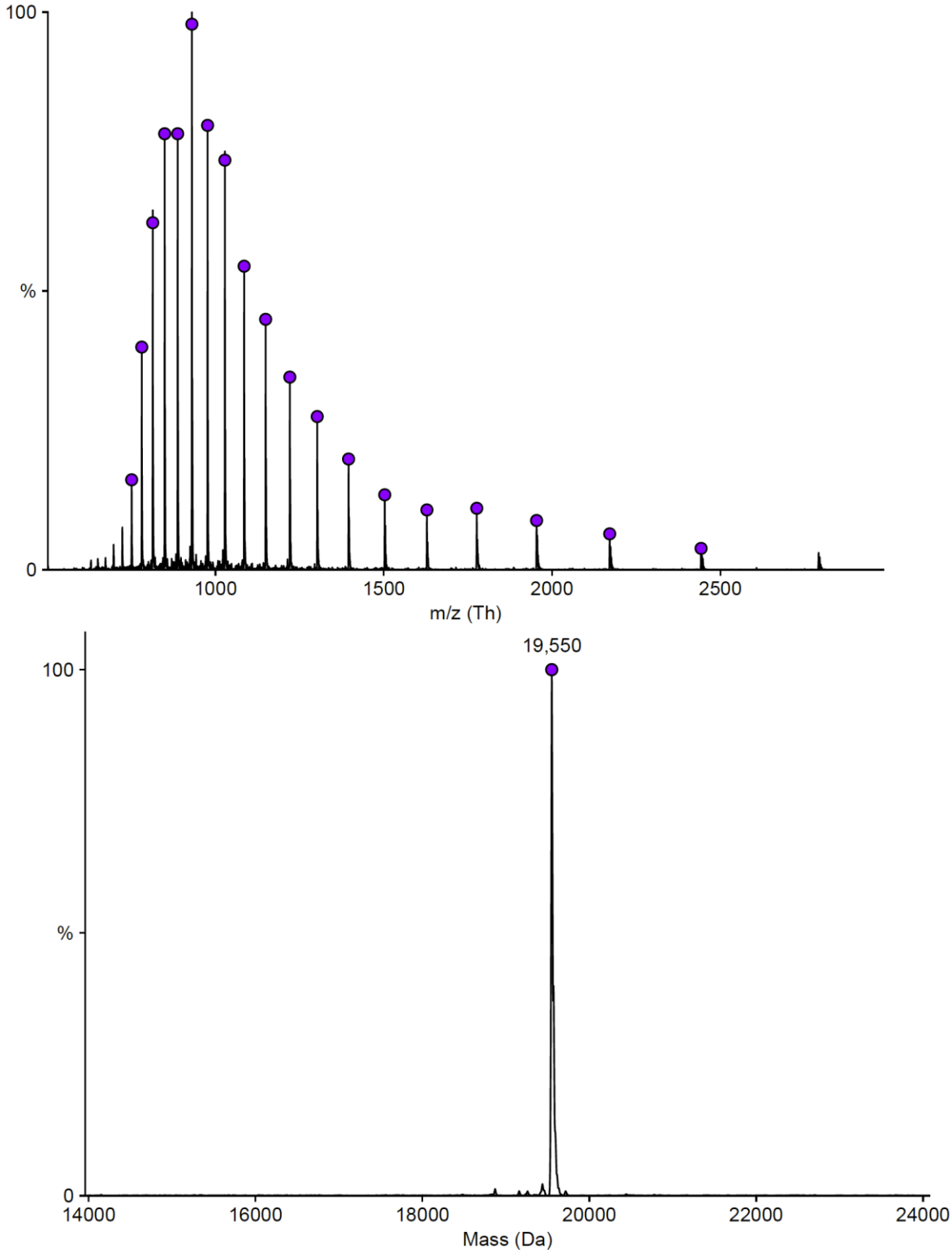


Figure S51. Raw and deconvoluted MS spectrum of 5´-Cy5-**HIR-8**, calculated mass: 19557 Da, found mass: 19550 Da.

13. References

1. Ondruš, M.; Sýkorová, V.; Bednárová, L.; Pohl, R.; Hocek, M.: "Enzymatic Synthesis of Hypermodified DNA Polymers for Sequence-Specific Display of Four Different Hydrophobic Groups" Nucleic Acids Res. **48**, 11982-11993 (2020).

2. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol*. **215**, 403–410 (1990).

3. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**, 10–12 (2011).

4. Aronesty, E. Comparison of sequencing utility programs. *Open Bioinform. J*. **7**, 1–8 (2013).

5. Gaspar, J. M. NGmerge: merging paired-end reads via novel empirically-derived models of sequencing errors. *BMC Bioinformatics* **19**, 536 (2018).

6. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* **11**, e0163962 (2016).

7 . Gordon, A. & Hannon, G. J. FASTX-Toolkit. *Hannon Lab* **http://hannonlab.cshl.edu/fastx_toolkit** (2014).

8. Kramer, S. T., Gruenke, P. R., Alam, K. K., Xu, D. & Burke, D. H. FASTAptameR 2.0: a web tool for combinatorial sequence selections. *Mol Ther Nucleic Acids* **29**, 862–70 (2022).

9 . Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

10. Kumar, S. et al. MEGA12: molecular evolutionary genetics analysis version 12 for adaptive and green computing. *Mol. Biol. Evol.* **41**, 1–9 (2024).

11. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

12. Špačková, B., Lynn, N. S., Slabý, J., Šípová, H. & Homola, J. A route to superior performance of a nanoplasmonic biosensor: consideration of both photonic and mass transport aspects. *ACS Photon*. **5**, 1019–1025 (2018).

13. Špringer, T., Piliarik, M. & Homola, J. Surface plasmon resonance sensor with dispersionless microfluidics for direct detection of nucleic acids at the low femtomole level. Sens. Actuators B Chem. **145**, 588–591 (2010).

14. Špringer, T., Chadtová Song, X., Ermini, M. L., Lamačová, J. & Homola, J. Functional gold nanoparticles for optical affinity biosensing. Anal. Bioanal. Chem. **409**, 4087–4097 (2017).

15. Morcavallo, A. et al. Insulin and insulin-like growth factor II differentially regulate endocytic sorting and stability of the insulin receptor isoform A. *J Biol Chem* **287**, 11422–11436 (2012).

16. Křížková, K. et al. The insulin-IGF hybrids as molecular probes of hormone: receptor binding specificity. *Biochemistry* **55**, 2903–2913 (2016).

17. Páníková, T. et al. Insulin analogues with altered insulin receptor isoform binding specificities and enhanced aggregation stabilities. *J Med Chem* **64**, 14848–14859 (2021).

18. Asai, S., Žáková, L., Selicharová, I., Marek, A. & J, J. A radioligand receptor binding assay for measuring of insulin secreted by MIN6 cells after stimulation with glucose, arginine, ornithine, dopamine, and serotonin. *Anal Bioanal Chem* **413**, 4531-4543 (2021).

19. Machackova, K. et al. Insulin-like Growth Factor 1 Analogs Clicked in the C Domain: Chemical Synthesis and Biological Activities. *J Med Chem* **60**, 10105–10117 (2017).

20. Lubos, M. et al. Modulation of the antagonistic properties of an insulin mimetic peptide by disulfide bridge modifications. *J Pept Sci* **29**, e3478 (2023).

21. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol*. **152**, 36–51 (2005).

22. Kimanius, D., Dong, L., Sharov, G., Nakane, T. & Scheres, S.H. W. New tools for automated cryo-EM single-particle analysis in relion-4.0. *Biochem. J*. **478**, 4169–4185 (2021).

23. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

24. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol*. **192**, 216–221 (2015).

25. McMullan, G., Vinothkumar, K. R. & Henderson, R. Thon rings from amorphous ice and implications of beam-induced Brownian motion in single particle electron cryo-microscopy. *Ultramicroscopy* **158**, 26–32 (2015).

26. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).

27. Croll, T. I. et al. Higher-resolution structure of the human insulin receptor ectodomain: multi-modal inclusion of the insert domain. *Structure* **24**, 469–476 (2016).

28. Burnley, T., Palmer, C. M. & Winn, M. Recent developments in the CCP-EM software suite. *Acta Crystallogr. D Struct. Biol*. **73**, 469–477 (2017).

29. Jakobi, A. J., Wilmanns, M. & Sachse, C. Model-based local density sharpening of cryo-EM maps. *Elife* **6**, e27131 (2017).

30. Naydenova, K. & Russo, C. J. Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nat. Commun*. **8**, 629 (2017).

31. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

32. Rosenthal, B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).

33. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25 (2010).

34. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132 (2004).

35. Long, F. et al. AceDRG: a stereochemical description generator for ligands. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 112–122 (2017).

36. Potterton, L. et al. CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallogr. D* **74**, 68–84 (2018).

37. Brown, A. et al. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D* **71**, 136–153 (2015).

38. Kovalevskiy, O., Nicholls, R. A., Long, F., Carlon, A. & Murshudov, G. N. Overview of refinement procedures within REFMAC5: utilizing data from different sources. *Acta Crystallogr. D Struct. Biol.* **74**, 215–227 (2018).

39. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).

40. Afonine, P. V. et al. Real-space refinement in phenix for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).

41. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D. Struct. Biol.* **75**, 861–877 (2019).

42. Williams, C. J. et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci*. **27**, 293–315 (2018).

43. Berman, H. M., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat Struct Biol* **10**, 980 10.1038/nsb1203-980 (2003).

44. Meng, E. C. et al. UCSF ChimeraX: tools for structure building and analysis. *Protein Sci*. **32**, e4792 (2023).

45. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr*. **66**, 486–501 (2010).

46. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol*. **372**, 774–797 (2007).

47. The PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC.