Identification and Rational Redesign of Peptide Ligands to CRIP1, A Novel Biomarker for Cancers

Jihua Hao¹, Adrian W.R. Serohijos^{4,5}, Gail Newton⁷, Gina Tassone⁷,
Zuncai Wang⁷, Dennis C. Sgroi⁷, Nikolay V. Dokholyan^{4,6,*}, James P. Basilion^{1, 2, 3,*}

¹Department of Radiology, ²Department of Biomedical Engineering, and ³National Foundation for Cancer Research Center for Molecular Imaging, Case Western Reserve University, Cleveland, OH 44106

⁴Department of Biochemistry and Biophysics, ⁵Department of Physics and Astronomy, ⁶Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

⁷ Harvard Medical School, Massachusetts General Hospital, Boston, MA 02129

Supplementary Information

Figure S1. The cDNA and amino acid sequences of CRIP1.

After cloning, the insert was confirmed by sequencing and the deduced amino acid sequences for the human CRIP1 shown in Figure S1. Excluding the vector sequence and the poly A region, the cDNA insert is 243 base pairs in length. The start site for transcription is at nucleotide position 73 (not shown in figure) with the start of translation at nucleotide position 162. This open reading frame expresses the amino acids encoding the His-tag (nt: 186-242) and encoding an enterokinase clevage site (nt: 246-260). The sequences encoding the human CRIP1 protein begin at nucleotide 273 and continue through nucleotide 503. Translation of these sequences results in a polypeptide 114 amino acids in length, the majority of which, 77 amino acids, make up CRIP1 protein.

The start (ATG) and stop (TAA) codons are underlined. The sequence of nonadjacent 6 histidines on HAT epitope is in bold. The poly A tail at the end is not shown.

								•		HindI	II					
CAG	GAA	ACA	GCT	M ATG	T ACC		_	_	_		L TTG	K AAG	D GAT	H CAT	CTC	I ATC
					Н	ΙΑΤ										
H CAC	N AAT	V GTC	H CAC	K AAA	E GAG	E GAG	H CAC	A GCT	H CAT	A GCC	H CAC	N AAC	K AAG	, I	D	D
			~	····	Ra	mHI										
D GAT			V	D	~~~~	~~~~ S	M ATG			C TGT	P	K AAG	C TGC	N AAC	K AAG	E GAG
			010	0.10	00.1	700	<u> </u>			101			100		1110	0.10
V GTG	Y TAC	F	A GCC	E GAG	R AGG	V GTG	ACC	S TCT	L CTG	G GGC	K AAG	D GAC	W TGG	H CAT	R CGG	P
C TGC	L CTG	K AAG	C TGC	E GAG	K AAA	C TGT	G GGG	K AAG	ACG	L CTG	T ACC	S TCT	G GGG	G GGC	H CAC	A GCT
E GAG			G GGC	K AAA	P	Y TAC	C TGC	N AAC	H CAC	PCCC	C TGC	Y TAC	A GCA	A GCC	M ATG	F TTT
G GGG	P CCT	K AAA	G GGC	F TTT	G GGG	R CGG	G GGC	G GGA	A GCC	E GAC	S G AGO	H CAC	T ACT	F TTC	K AAG	* TAA
*	~~~	····		~~~~	~ ~~	CCT	TTT									
	H CAC D GAT EK c V GTG C TGC E GAG GGGG *	H N CAC AAT D D GAT GAC EK cleavage V TAC C CTG E GAG CAC GGG CCT E CCT E CCT GGG CCT E CC	H N V V O CAC AAA O T O CAC AAA O CAC AAA O CAC CTG AAAA O CAC CTG AAA A	H N V CAC H CAC AAT GTC CAC H S S S S D D D K V V GAT GAC AAA GTC EK cleavage site V T Y F GAC C L K TTC GCC C CTG AAG TGC GAG CAC GAA GGC GGG P K GGC GGGG CCT AAA GGC EcoRI * **********************************	H N V H K CAC AAT GTC CAC AAA Bincil Sall Sall V D D K GAT GAC AAA EK cleavage site V T T GC GAC C C L K C GAC TGC CTG AAG C C CTG AAG GGG CAC GAA GGG CTTT EcoRI Eael Eael C ATG ATG ATG ATG ATG ATG ATG ATG	CAG GAA ACA GCT ATG ACC H H CAC AAT GTC CAC AAA GAG CAC AAA GTC Ball Ba W GAT GAC AAA GTC EK cleavage site CC CT CT CT CT CT CT CT CT	CAG GAA ACA GCT ATG ACC ATG HAT H N V H K E E E CAC AAA GAG GAG HincII Sall BamHI Sall BamHI O D G S GAT GAC AAA GTC GAC GAC GGA TCC EK cleavage site V T T GAC GAC GAC AAA CC TGC CTG AAG TGC GAC GAG AAA CC TGC CTG AAG GGC AAA CCC TAC GGG CCT AAA GGC TTT GGG CGG ECORI * HAT ATG HAT ACC ATG ATG ACC ATG HINCII Sall BamHI ACC AAA GAG GAG AAA TCC AAA AAA GTC GAC GAA ACC AAA CC TGC GAG AAA CC TAC CG GGG CT AAA GGC TTT GGG CGG C	CAG GAA ACA GCT ATG ACC ATG ATT HAT H N V CAC AAA GAG GAG CAC Hincll Sall BamHI Sall BamHI FINAT BamHI SALL BAT BAT BAT BAT BAT BAT BAT B	CAG GAA ACA GCT ATG ACC ATG ATT ACG HAT H N V H K E E E H A CAC AAT GTC CAC AAA GAG GAG CAC GCT HincII Sall BamHI SALC ATG CCC EK cleavage site V D G S M P GAT GAC AAA GTC GAC GAG AGG CCC EK cleavage site V T T T S GTG TAC TTC GCC GAG AGG GTG ACC TCT C L K C C C GAG AAA GTC GAC AAA TGT GGG AAG GGG CCT AAA GGC TTT GGG CGG GGC GGAA EcoRI Eael **	CAG GAA ACA GCT ATG ACC ATG ATT ACG CCA HAT H N V CAC AAA GTC CAC AAA GAG GAG CAC GCT CAT Hinell Sall BamHI O D D K V D G S M P K GAT GAC AAA GTC GAC GGA TCC ATG EK cleavage site V D D G S M P K GAT GAC AAA GTC GAC GGA TCC ATG CC CT AAA TCC GAC GAG AGG GTG ACC TCT CTG CC CTG AAG TGC GAG AAA TGT GGG AAG ACG GGAG CAC GAA GGC AAA CCC TAC TGC AAC CAC GGG CCT AAA GGC TTT GGG CGG GGC GGA GCC ECORI Eael ** ** ** ** ** ** ** ** **	CAG GAA ACA GCT ATG ACC ATG ATT ACG CCA AGC HAT HAT CAC AAT GTC CAC AAA GAG GAG GAG CAC GCT CAT GCC Hincli Sall BamHI Sall BamHI CAC GAT GAC AAA GTC GAC GGA TCC ATG CCC AAG TGT EK cleavage site GTC GAC GAG AGG GTG ACC TCT CTG GGC CC L K C CTG AAG TGC GAG AAA TGT GGG AAG ACC CTG GAG CAC GAG AAA TGT GGG AAG ACC CTG GAG CAC GAG AAA TGT GGG AAG ACC CTG GAG CAC GAG AAA TGT GGG AAC CAC CCC GGG CAC GAA GGC TTT GGG CAC GAG AAC CAC CCC GGG CCT AAA GGC TTT GGG CAC GAG AAC CAC CCC GGG CCT AAA GGC TTT GGG CAC GGA GGC GGA GCC GAC ECORI Eael ** ** ** ** ** ** ** ** **	CAG GAA ACA GCT ATG ACC ATG ATT ACG CCA AGC TTG HAT HAT H N V H K E E E H A H A H A H A H A H A H A H A A H A A H A	CAG GAA ACA GCT ATG ACC ATG ATT ACG CCA AGC TTG AAG HAT HAT HAT CAC AAT GTC CAC AAA GAG GAG GAG CAC GCT CAT GCC CAC AAC Hincli Sall BamHI Sall BamHI SAG CCC AAG TGT CCC AAG EK cleavage site CC CTG AAA TGC GAC GAG AAA TGT GGG AAG TGT CCC AAG GTG TAC TTC GCC GAG AAA TGT GGG AAG ACC CTG CTG AAA GGC TTT GGG CAG AAA TGT GGG AAG ACC CTG CTG AAA GGC TTT GGG CAG AAA TGT GGG AAA TGT GGG AAA TGT GGG CAC GAA GGC TTT GGG CAG GGA GCC GAG AGC CAC CTG CTG AAA GGC TTT GGG CAG GGA GCC GAG AGC CAC CTG CTG AAA GGC TTT GGG CAG GGA GCC GAG AGC CAC CTG CTG AAA GGC TTT GGG CAG GGC GGA GCC GAG AGC CAC CTG CTG AAA GGC TTT GGG CAG GGC GGA GCC GAG AGC CAC CTG CTG AAA GGC TTT GGG CAG GGC GGA GCC GAG AGC CAC CTG CTG AAA GGC TTT GGG CAG GGC GGA GCC GAG AGC CAC CTG CTG AAA GGC TTT GGG CAG GGC GGA GCC GAG AGC CAC CTG CTG AAA GGC CTG AACC CAC CTG CTG AAA GGC CTG CAC CTG CTG CTG CTG CAC CTG CTG CTG CTG CAC CTG CTG C	CAG GAA ACA GCT ATG ACC ATG ATT ACG CCA AGC TTG AAG GAT HAT HAT HINCH Sall BamHI Sall BamHI Sall BamHI SAT GAC AAG GCT AAG GCC AAG TGT CCC AAG TGC EK cleavage site C CAG GAA AAG GAG GTG ACC TCT CTG GGC AAG GAC TGC TGC CTG AAG TGC GAC AAA TGT GGG AAA TGT GGG AAG AAG AAG AAG	CAG GAA ACA GCT ATG ACC ATG ATT ACG CCA AGC TTG AAG GAT CAT HAT HINCILL Sall BamHI SAL AAG GAC AAG GAC AAG GAC AAG GAC AAG AAG	CAG GAA ACA GCT ATG ACC ATG ATT ACG CCA AGC TTG AAG GAT CAT CTC HAT CAC AAT GTC CAC AAA GAG GAG GAG CAC ATG ATT ACG CCA AGC TTG AAG GAT CAT CTC HAT CAC AAT GTC CAC AAA GAG GAG GAG CAC GCT CAT GCC CAC AAC ACG ATC GAT HincII Sall BamHI BAMHI SALL BAMHI BAMHI SALL BAMHI SALL BAMHI BAMHI SALL BAMHI SALL BAMHI SALL BAMHI BAMHI SALL BAMHI BAMHI SALL BAMHI SALL BAMHI BAMHI SALL BAMHI BAMHI SALL BAMHI BAMHI SALL BAMHI SALL BAMHI BAMHI SALL BAMHI BAMHI SALL BAMHI BAMHI SALL BAMHI BAMHI BAMHI SALL BAMHI BAMHI BAMHI SALL BAMHI BAMHI SALL BAMHI BAMHI

Figure S2. CRIP1 Purity. Comassie Blue Stained-SDS-PAGE analysis of CRIP1 lysate and elutions after purification.

Lane 1: standard molecular marker;

Lane 2: lysate before incubation with Resin;

Lane 3: lysate after incubation with Resin; Lane 4~5: fractions through Clontech TALON CellThru column;

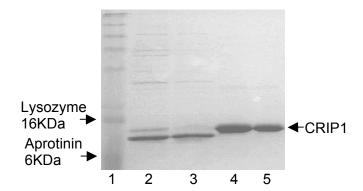


Figure S3. Relative estimates of peptide affinity for CRIP1. Phage binding against immobilized CRIP-1.

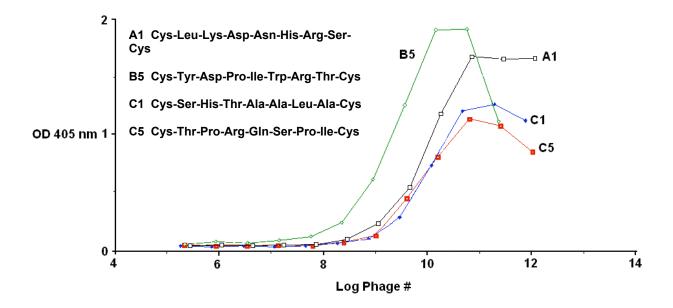


Figure S4. Synthesis of FITC-peptides

Boc-Cys-Leu-Lys-Asp-Asn-His-Arg-Ser-Cys-Gly-Gly-Gly-Ser-Lys(Dde)-Resin

Figure S5. Sequences of redesigned peptides. Shown below are the sequence motifs of the redesigned peptides for the starting peptide structure models 1-ns, 9-ns, and 10-ns. The rank pertains to the order putative the binding site on CRIP1 defined from clustering.

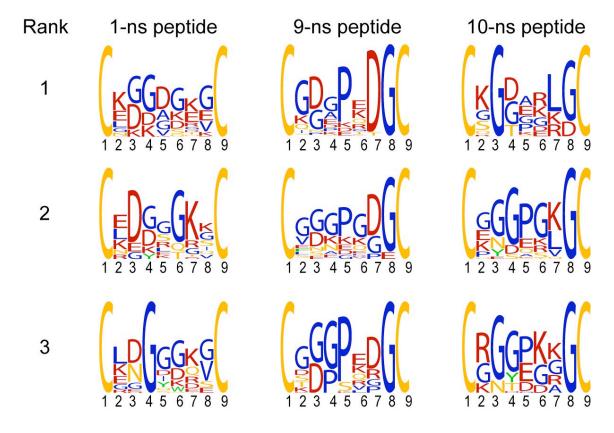


Figure S6. Residue preference without CRIP context. To verify that the observed preference for Gly in some sites in the peptide is not due to a bias in the force field, we employed the protocol to find the optimal peptide sequence when the peptide is not bound to CRIP1. We used 50 independent redesign runs. The preferred sequences are expectedly highly polar which maximize the peptide solvation energy.



6

Table S1. Analysis of CRIP1

Analysis	Entire Protein
Length	114 aa
Molecular Weight	12743.72
1 microgram	78.470pMoles
Molar Extinction	10370
coefficient	
1 A[280] corr. To	1.23 mg/ml
A[280] of 1 mg/ml	0.81 AU
Isoelectric point	8.20
Charge at pH 7	2.66

Table S2. Contribution of individual energy terms to the $\Delta\Delta G$ of the redesigned peptide A1M CLDGGGKGC.

Energy Term ^a	Symbol	Value (kcal/mol)
van der Waals attraction	evdw_a	17.3
van der Waals repulsion	evdw_r	-99.1
Solvation	esolv	-15.6
Hydrogen bond interactions between sidechain and backbone	ehb_sb	1.1
Hydrogen bond interactions between sideschains	ehb_ss	0.0
Internal energy for a rotamer state	efy_chi	1.0
Correction for dependence rotamer energy on backbone conformation and amino acid type	efy_aa	2.7
Reference energy ^b	e_aa_ref	10.0
Total ΔΔG		-83

^aPlease see Ding and Dokholyan³⁹ for detailed calculation of the force field terms.

bReference energy assumes that the polypeptide is completely unfolded.