

Supplementary Information for

Identification of a biologically active fragment of ALK and LTK-Ligand 2 (Augmentor-α)

Andrey V. Reshetnyak^{1*}, Jyotidarsini Mohanty¹, Francisco Tomé¹, David E. Puleo¹, Alexander Plotnikov¹, Mansoor Ahmed¹,Navjot Kaur², Anton Poliakov³, Arul M. Cinnaiyan³, Irit Lax¹, and Joseph Schlessinger¹

Corresponding authors Joseph Schlessinger and Andrey Reshetnyak Email: <u>joseph.schlessinger@yale.edu</u> and <u>andrey.reshetnyak@stjude.org</u>

This PDF file includes:

Supplementary text Figs. S1 to S3

Supplementary Information Text

Protein sequences of Fc-AUG- α (mammalian expression construct cloned in pCEP4 vector):

METDTLLLWVLLLWVPGSTGAGSTHTCPPCPAPELLGGPSVFLFPPKPKDTLMI SRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSV LTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELT KNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTV DKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGKASGAGSTTGGGGSGGG GSGGGGS**LEVLFQ/GP**GAEPREPADGQALLRLVVELVQELRKHHSAEHKGLQL LGRDCALGRAEAAGLGPSPEQRVEIVPRDLRMKDKFLKHLTGPLYFSPKCSKH FHRLYHNTRDCTIPAYYKRCARLLTRLAVSPVCMEDKQ

Signal peptide is in the italic; 3C protease cleavage site is in bold (/ is the actual cleavage site)

Protein sequences of Fc-ALK (mammalian expression construct cloned in pCEP4 vector):

METDTLLLWVLLLWVPGSTGAGSTHTCPPCPAPELLGGPSVFLFPPKPKDTLMI SRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSV LTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELT KNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTV DKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGKTSLYKKAGF**ENLYFQ/GL** <u>NDIFEAQKIEWHE</u>TAPKSRNLFERNPNKELKPGENSPRQTPIFDPTVHWLFTTC GASGPHGPTQAQCNNAYQNSNLSVEVGSEGPLKGIQIWKVPATDTYSISGYGA AGGKGGKNTMMRSHGVSVLGIFNLEKDDMLYILVGQQGEDACPSTNQLIQKVC IGENNVIEEEIRVNRSVHEWAGGGGGGGGGGGTYVFKMKDGVPVPLIIAAGGGGR AYGAKTDTFHPERLENNSSVLGLNGNSGAAGGGGGGWNDNTSLLWAGKSLQE GATGGHSCPQAMKKWGWETRGGFGGGGGGGCSSGGGGGGYIGGNAASNND PEMDGEDGVSFISPLGILYTPALKVMEGHGEVNIKHYLNCSHCEVDECHMDPE SHKVICFCDHGTVLAEDGVSCIVSPTPEPH<u>GLNDIFEAQKIEWHE</u>

Signal peptide is italic; TEV protease cleavage site is in bold; Biotin Acceptor Peptide is underlined.

Protein sequences of Trx-AUG-α AD *E. coli* expression construct cloned in pET42 vector):

MHHHHHHENLYFQ/GSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIA PILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALS KGQLKEFLDANLAGGGGSLEVLFQ/GPSPEQRVEIVPRDLRMKDKFLKHLTGPL YFSPKCSKHFHRLYHNTRDCTIPAYYKRCARLLTRLAVSPVCMEDKQ

TEV protease and 3C protease cleavage sites are in bold (/ is the actual cleavage site);

Fig. S1 Multiple Sequence Alignment.

Δ				
~	Human Chimpanzee Mouse Rat Chicken GiantPanda	MRGPGHPLLLGLLVLGAAGRGRGGAEPREPADGQALLRLVVELVQELRKHHSAEHKGLQLLGR-D MGGPGRPLLLGLLVLGAAGRGRGGAEPREPADGQALLRLVVELVQELRKHHSAEHKGLQLLGR-D MRVSGRPMLLALLLLSTVGDRGRAQSRGPADRQTLLRLVELVQELKKFHIGDSKRLQLLGESD MRVSGRPMLLALLLLSTVGDPGHAQPRGPADRQTLLRLVELVQELKKFHIGDSKRLQLLGESD MSGLRSPGLLGVLLVLSAGYCKEKTDAADLKDRQSLLNLIMEIIQELKRYRLEKDSGVQYFSKHD MRGPGRPLLLGLLVLGAAGPGVAEPREAADRQTLLRLVEILQELKKYHSGESKRLPLSGQHD * * ***.*:::::* : * * *:*:*::***	CALGRAEAAGLGPSP- CALGRAEAAGLGPSP- FALGRREATDYGADQE FALGRREATDYGADQE YSLDRREVADYGGYQD YTLGRREASDYPVYPE	80 80 81 81 82 80
	Human Chimpanzee Mouse Rat Chicken GiantPanda	EQRVEIVPRDLRMKDKFLKHLTGPLYFSPKCSKHFHRLYHNTRDCTIPAYYKRCARLLTRLAVSPVCMED EQRVEIVPRDLRMKDKFLKHLTGPLYFSPKCSKHFHRLYHNTRDCTIPAYYKRCARLLTRLAVSPVCMED EQRVEIVPRDLRMKDKFLKHLTGPLYFSPKCSKHFHRLYHNTRDCTIPAYYKRCARLLTRLAVSPMCMEB EQRVEIVPRDLRMKDKFLKHLTGPLYFSPKCSKHFHRLYHNTRDCTIPAYYKRCARLLTRLAVSPMCMEB EQRVEIVPRDLRMKDKFLKHLTGPLYFSPKCSKHFHRLYHNTRDCTIPAYYKRCARLLTRLAVSPMCMEB EQRVEIVPRDLRMKDKFLKHLTGPLYFSPKCSKHFHRLYHNTRDCTIPAYYKRCARLLTRLAVSPMCMEB	KQ 152 K 151 151 151 152 150	
В				
	Human Mouse Rat Rabbit Horse GiantPanda	MRPLKPGAPLPALFLLALALSPHGAHGRPRGRRGARVTDKEPKPLLFLPAAGAGRTPSG59MWLTKPSTPVSALLLLALALSPPGTQGRPQRSLAARVAELRPELFLPVTGTRLPPRA57MWLAKLSTRVLLLLLALALSPPGTQGRPGQRRGARARVTELLPQLFLPVTGTRLAPRT57MRPAKPGAPLPALLLTLALSPHGTQGRPQGRRGARVAGEEPKTLLFVPAAGAGPAPSA59MRPAKPGATLPALLLLALVLAPHETQGRPRGRRGARVANEEPKPSLFLPAAGAGPASSA59MRLAKPGAPLPALLLVLALALSLPGAQGRPRGGRGARATVQESKSWLFLPAARAGPAPHA59***		
	Human Mouse Rat Rabbit Horse GiantPanda	SRSAEIFPRDSNLKDKFIKHFTGPVTFSPECSKHFHRLYYNTRECSTPAYYKRCARLLTRLAVSPLCSQT SRSTEIFPRDLTLKDKFIKHFTGPVTFSAECSKHFHRLYHNTRDCSTPAYYKRCARLLTRLAVSPLCSQT SRSTEIFPRDLTLKDKFIKHFTGPVTFSAECSKHFHRLYHNTRDCSMPAYYKRCARLLTRLAVSPLCSQT SRSAEIFPRDLNLKDKFIKHFTGPVTFSAECSKHFHRLYHNTRDCSTPAYYKRCARLLTRLAVSPLCSQT SRSAEMFPRDLNLKDKFIKHFTGPVTFSAECSKHFHRLYHNTRDCSTPAYYKRCARLLTRLAVSPLCSQT SRSAEMFPRDLNLKDKFIKHFTGPVTFSAECSKHFHRLYHNTRDCSTPAYYKRCARLLTRLAVSPLCSQT SRSAEMFPRDLNLKDKFIKHFTGPVTFSAECSKHFHRLYHNTRDCSTPAYYKRCARLLTRLAVSPLCSQT SRSAEMFPRDLNLKDKFIKHFTGPVTFSAECSKHFHRLYHNTRDCSTPAYYKRCARLLTRLAVSPLCSQT SRNTEIFPRDLNLKDKFIKHFTGPVTFSAECSKHFHRLYHNTRDCSTPAYYKRCARLLTRLAVSPLCSQT	129 127 127 129 129 129	
С				
	AUG-a EQ	RVEIVPRDLRMKDKFLKHLTGPLYFSPKCSKHFHRLYHNTRDCTIPAYYKRCARLLTRLAVSPVCMEDKQ	152	
	AUG-B SR	SAEIFPRDSNLKDKFIKHFTGPVTFSPECSKHFHRLYYNTRECSTPAYYKRCARLLTRLAVSPLCSOT	129	

Multiple sequence alignment of AUG- α (A) and AUG- β (B) from different species. Signal peptide is highlighted in yellow, N-terminal variable region in green, and C-terminal conserved augmentor domain in blue. All cysteines are marked with red. (C) Sequence alignment of conserved augmentor domain of human AUG- α and human AUG- β .

Α **GP**GAEPREPADGQALLRLVVELVQELRKHHSAEHKGLQLLGRDCALGRAEAAGLGPSPEQRVEIVP RDLRMKDKFLKHLTGPLYFSPKCSKHFHRLYHNTRDCTIPAYYKRCARLLTRLAVSPVCMEDKQ **GP**GAEPREPADGQALLRLVVELVQELRKHHSAEHKGLQLLGRDCALGRAEAAGLGPSPEQRVEIVP RDLRMKDKFLKHLTGPLYFSPKCSKHFHRLYHNTRDCTIPAYYKRCARLLTRLAVSPVCMEDKQ Β 29402 6.5e4 Charge Thepritical avarage MW, Da MH1+ 29400.763 6.0e4 MH18+ 1634.32 MH19+ 1548 362 5.5e4 MH20+ 1470.995 MH21+ 1400.995 5.0e4 MH22+ 1337 359 MH23+ 1279.257 4.5e4 MH24+ 1225,997 4.0e4 MH25+ 1176.997 MH26+ 1131.767 3.5e4 MH27+ 1089.887 3.0e4 2.5e4 2.0e4 15155 1.5e4 29417 21880 1.0e4 14701 5000.0 29513 0371 20892 22622 23954 17614 15007 20195 14156 5958 26955 18251 11006 13383 24289 28349 29628 29905 24601 44 0.0 1.0e4 بلاب ا. م الله الله ال 1.2 e4 1.4e4 1.6 e4 1.8e4 2.0 e4 2.2e4 2.4e4 2.6e4 2.8e4 3.0e4 Mass, Da С 1226 1279 1177 9000 8500 1337 8000 7500 7000 1132 6500 6000 14,01 5500 1090 5000 4500 892 4000 1051 843 3500 1471 3000 1015 2500 83 95 1153 2000 1500 1548 1000 1634 500 41_1656 1731_1750 569 1700 1800 1200 1400 1500 1600 1000 1100 1300 900

m/z. Da

Fig. S2 Mass Spectrometric Analysis of Dimeric AUG- α .

(A) Primary structure of dimeric AUG-α. First two residues highlighted in bold are artificial residues from 3C protease cleavage site. All cysteines are highlighted in red and bold. Intramolecular disulfide bridges are shown with solid red lines; intermolecular disulfide bridge is shown with dash red line.

(B and C) ESI-TOF high accuracy mass spectrometric analysis of dimeric AUG-α. Theoretically predicted average molecular mass is represented in the insert of(B).



(A and B) SDS-PAGE analysis of ALK fragment (residues 648-1025) after limited proteolysis with a set of proteolytic enzymes. ALK fragment was incubated with different proteolytic enzymes at a ratio 1:100 protease:ALK at room temperature for 1 hour, and cleavage reaction was monitored by SDS-PAGE. Abbreviations are as follows: $\alpha C - \alpha$ -Chymotrypsin, TR – trypsin, EL – elastase, PA – papain, SU – subtilisin and endoproteinase Glu-C. - is a negative control ALK fragment which was not incubated with any proteolytic enzymes. (B) Aug- α AD was added into limited proteolysis reaction where indicated (+). The ratio of protease:ALK was increased to 1:30, and ALK:Aug- α AD ratio was 1:5.

(C) Bands which were used for N-terminal sequencing are marked by arrows and corresponding sequences are shown for each band.

(D) Sequence alignment of putative ALK TNF-like motif with TRAIL motif. Alignment was performed using the Phyre2 server [1]. (E) Schematic representation of ALK and LTK domain organization. N-terminal domain (NTR) colored in cyan, MAM in green, LDL in yellow, TNF-like motif in dark blue, glycine-rich (Gly-Rich) in orange, EGF-like motif in purple, transmembrane (TM) in black, and kinase domain in blue. LTK domains are shown with striations.

References

1. Kelley LA et al. (2015) The Phyre2 web portal for protein modeling, prediction and analysis *Nature Protocols* 10: 845-858.