Cellular and Molecular Life Sciences (2019) 76:5027–5039
https://doi.org/10.1007/s00018-019-03159-5

ORIGINAL ARTICLE

Open reading frame mining identifies a TLR4 binding domain in the primary sequence of ECRG4

Xitong Dang[1,2] · Raul Coimbra[1] · Liang Mao[2] · Sonia Podvin[1] · Xue Li[2] · Hua Yu[2] · Todd W. Costantini[1] · Xiaorong Zeng[2] · Dana Larocca[3] · Brian P. Eliceiri[1] · Andrew Baird[1,4]

Received: 22 January 2019 / Revised: 29 April 2019 / Accepted: 22 May 2019 / Published online: 12 June 2019

Abstract

The embedding of small peptide ligands within large inactive pre-pro-precursor proteins encoded by orphan open reading frames (ORFs) makes them difficult to identify and study. To address this problem, we generated oligonucleotide (< 100–400 base pair) combinatorial libraries from either the epidermal growth factor (EGF) ORF that encodes the > 1200 amino acid EGF precursor protein or the orphan ECRG4 ORF, that encodes a 148 amino acid Esophageal Cancer Related Gene 4 (ECRG4), a putative cytokine precursor protein of up to eight ligands. After phage display and 3–4 rounds of biopanning for phage internalization into prostate cancer epithelial cells, sequencing identified the 53-amino acid EGF ligand encoded by the 5′ region of the EGF ORF and three distinct domains within the primary sequence of ECRG4: its membrane targeting hydrophobic signal peptide, an unanticipated amino terminus domain at ECRG4$^{37–63}$ and a C-terminus ECRG4$^{133–148}$ domain. Using HEK-blue cells transfected with the innate immunity receptor complex, we show that both ECRG4$^{37–63}$ and ECRG4$^{133–148}$ enter cells by interaction with the TLR4 immune complex but neither stimulate NFkB. Taken together, the results help establish that phage display can be used to identify cryptic domains within ORFs of the human secretome and identify a novel TLR4-targeted internalization domain in the amino terminus of ECRG4 that may contribute to its effects on cell migration, immune cell activation and tumor suppression.

Keywords Ligand · Receptor · Peptide targeting · Secretome · Phage display · Cytokine precursor · Tumor suppressor gene

Introduction

Bioinformatic analyses of the human genome initially revealed that remarkably few genes, then estimated at approximately 25,000, were present given the complexity of the human proteome [1]. Although these findings can be reconciled by the transcriptional and translational complexity that is introduced by both alternative splicing of RNA [2, 3] and the post-translational processing of proteins [4, 5], it has underscored the need to identify active but cryptic protein products encoded within open reading frames. Accordingly, there are likely to be numerous peptide ligands that are overlooked because they are hidden within high molecular weight protein precursors [6, 7].

Fortunately, precursor proteins contain consensus sequences that predict protein secretion and post-translational processing. In this way, bioinformatic algorithms using techniques like Markov Modeling [8–10] can predict the existence of cryptic ligands in the secretome [11, 12]. That being said, there are few experimental approaches that can ascribe biological functions to peptides embedded in open reading frames (ORFs).

The 148 amino acid esophageal cancer related gene 4 (ECRG4) protein is a case in point. Encoded by the orphan C2orf40 gene, its differential gene expression in esophageal and other cancers [13, 14] first suggested its identity as a tumor suppressor and indeed, it has been linked to cancer development [15–22] tumor progression and metastases [23–25]. Examination of its protein sequence, however,

✉ Andrew Baird
abaird@ucsd.edu

1 Department of Surgery, University of California San Diego, San Diego, CA 92103, USA

2 The Key Laboratory of Medical Electrophysiology, Ministry of Education, Institute of Cardiovascular Research, Southwest Medical University, Luzhou 646000, China

3 Mandala Biosciences, San Diego, CA, USA

4 Department of Surgery, University of California San Diego, La Jolla, San Diego, CA 98896, USA

reveals a neuropeptide-like pre-pro-hormone motif [26] that is reminiscent of secreted proteins and significantly different from the structural features that are normally found in canonical tumor suppressor genes. Likewise, its physiological expression is tied to normal fetal development [27], CNS responsiveness and neuronal progenitor cell fate after injury [28], neuronal function [29], apoptosis [30], chondrocyte differentiation [31], cell senescence [32], lung preconditioning [33], wound healing [34], the infection response of epithelial mucosa [35], neural progenitor and, stem cell migration and differentiation [27, 36, 37], inflammation and leukocyte migration [24, 38–40], to name a few.

It is in light of its primary precursor-like sequence and its disparate activities, we [39] and others [41, 42] hypothesized that the wide distribution of ECRG4 in normal tissues might reflect different activities in different tissues. Depending on its selective processing and release from the cell surface [43, 44], the epigenetic regulation of its promoter [45] and the presence of receptors on effector cells, ECRG4 could have a myriad of activities. Accordingly, we reasoned that there is a need to identify the potential products of ECRG4 processing and their cognate binding partners on target cells.

In the work described here, we combined phage display of a fragmented ECRG4 open reading frame with biopanning on PC3 epithelial cells to identify internalizing domains. In addition to the hydrophobic signal sequence at the amino terminus ($ECRG4^{1–31}$) that is responsible for placement of ECRG4 at the plasma membrane cell surface [44] and a C-terminal $ECRG4^{133–148}$ domain that had been previously shown to bind to the LOX-1 scavenger receptor [19] and activate NFkB [38] a third domain was detected that encodes a new and unanticipated amino terminus domain that interacts with TLR4 to enter cells.

Methods

Materials

All chemicals, media, peptides and proteins were purchased from Sigma Chemicals (St Louis MO) unless specified otherwise. Recombinant $ECRG4^{31–70}$ was expressed in BL21DE3pLysS cells using the multicloning site of the commercial pET15b expression plasmid (Novagen EMD). Soluble protein was purified to homogeneity using FPLC and exploiting the amino terminus HIS-Tag of the recombinant protein before thrombin cleavage. Purity (> 90%) was assessed by SDS PAGE and shown to be free of endotoxin by LAL chromogenic quantification (ThermoFisher).

Tissue culture

The human PC-3 epithelial prostate cancer cell line and human HEK293 cell line (ATCC, Manassas, VA) and human dermal fibroblasts (ATCC, Manassas, VA) were maintained in RPMI1640 and HDMEM, respectively, supplemented with the additives described below. TLR4-HEK-blue cells that report innate immunity receptor activation of NFκb (see below) were purchased from InvivoGen (San Diego CA) and propagated as recommended by the manufacturer. All media were supplemented with 1 mM sodium pyruvate, non-essential amino acids, 50 units/50 µg/ml of penicillin/streptomycin (Invitrogen, Carlsbad, CA) and 10% fetal bovine serum.

Plasmid construction

Plasmid pcDNA3.1-hECRG4 was constructed as previously described (Dang et al. 2012). A pcDNA3.1-hECRG4$^{\Delta 31–70}$ plasmid was designed to encode an ECRG4 analog we call $ECRG4^{\Delta 31–70}$ that has amino acids 31–70 deleted from the parent $ECRG4^{1–148}$ sequence. Overlapping PCR used pcDNA3.1-hECRG4 as the template and the following primers to create the mutant:

5′-outer: ATACGGATCCATGGCTCCCCCGCG-3′,
3′-outer: 5′- ATACGAATTCTTAAGCGTAATCTGGAACATC-3′,
Del-up: 5′- CCAGGTGGCATAAGTCAGCTGTGGGACCGG,
Del-down: 5′- CCGGTCCCACAGCTGACTTATGCCACCTGG-3′.

PCR products were cloned into pcDNA3.1 at BamHI and EcoRI restriction sites and the purified plasmid sequenced to confirm the in-frame deletion of $ECRG4^{31–70}$ amino acids.

ORF fragmentation and phage display

The human EGF ORF was amplified by PCR with sense (5′-ATGCTGCTCACTCTTATC-3′) and anti-sense (5′-TCACTGAGTCAGCTCCATTT-3′) primers spanning start (ATG) to stop codons (TGA). The original template containing full-length human EGF cDNA was purchased from Open Biosystems (Huntsville, AL) and the PCR conditions set to 95 °C for 3 min, followed by 35 cycles of 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 30 s, and a final step of 72 °C for 5 min. The final PCR products were purified using a PCR purification kit (Qiagen, Valencia, CA) and quantified. Multiple samples of one µg PCR products were then partially digested with DNase I (1:500 dilution) in a volume of 20 µl at 16 °C for 15 min. Partially fragmented products

were resolved on 2% agarose gel, and fragments ranging from 200 to 400 bp were gel purified. The ends of the fragments were then blunted with T4 DNA Polymerase and Klenow fragments at room temperature for 30 min in 1 × T4 DNA Polymerase buffer (New England Biolabs, Ipswitch, MA) supplemented with 100 nM dNTPs. Adapters, annealed from XmnI C strand (5′-pCGAACCCCTTCG-3′) and NotI-XmnI non-palindromic (5′-GGCCCGAAGGGGTTCG-3′) adapters (NEB, Ipswich, MA), were ligated to the ends of the blunted fragments by T4 DNA ligase and incubated at 16 °C overnight. The adapter fragments were phosphorylated by T4 Polynucleotide kinase at 37 °C for 30 min. Using standard molecular cloning techniques, the fragments were then cloned into the pUC250 phagemid vector.

The fragmented human ECRG4 ORF library was prepared in the same way as human EGF ORF library with minor differences. Briefly, the human ECRG4 ORF was amplified by PCR with sense (5′- ATGGCTGCCTCCCCC GCGCGGC-3′) and antisense (5′-TTAGTAGACATCGTC GTTGACGC-3′) primers on pCMV6-XL4 that contained the full-length C2orf40 cDNA (Origene, Rockville, MD). The PCR conditions were the same as that for human EGF ORF amplification. The PCR product was purified and partially digested with DNase-I as described above. Partially fragmented products were then resolved on 2% agarose gel, and because the ORF is smaller than EGF, fragments ranging from 100–200 bp were gel purified. The ends of the fragments were then blunted with T4 DNA Polymerase and Klenow fragments at room temperature for 30 min. Adapters, annealed from two oligos 5′-p-CGGGCCGGCCGG CCGGCTGCA-3′ and 5′-GCCGGCCGGCCGGCCCG-3′ that carries the restriction site Pst-I, were added to the above blunted fragments by incubation with T4 DNA ligase at 16 °C for overnight. The adapter fragments were purified and cloned into a pUC198 phagemid vector [46]. To construct fragmented ECRG4 ORF phage library, less than 100 bp random fragments were gel-purified, blunted with T4 DNA Polymerase and Klenow fragments, and adapter, annealed from 5′-p-CGGCCGACTGT-3′ and 5′-ACAGTC GGCCG-3′ that carries an EagI restriction site, was added by T4 DNA ligase. The adapter fragments were then purified and cloned into the M13KE phage vector (NEB). The quality of the fragmentation libraries was evaluated by randomly selecting and digesting colonies with appropriate restriction enzymes and the ratio of colonies with and without inserts was calculated.

Phage biopanning on PC3 cells

Three to four rounds of biopanning were performed on PC-3 cells for recovery of internalized targeted puc198 phage. For each round of biopanning, PC-3 cells were seeded at 1×10^5 per well in a 12-well plate on the day before biopanning and the next day, cells were incubated with 1×10^{10}PFU phage particles for 4 h, sequentially washed three times with PBS containing 0.1% BSA and 50 mM Glycine, pH 2.8, respectively. Cells were then lysed with 0.5 ml lysis buffer (30 mM Tris–HCl, pH 8.0/0.1% Triton X-100) and sonicated for 5–10 s. Phage from the lysate were recovered and about 1×10^4 PFU were used for phage amplification. The amplified phage were used for the subsequent round of biopanning. After 3 or 4 rounds of biopanning, internalized phage in the lysate were recovered and phage DNA from well-isolated plaques was prepared and inserts were sequenced with a primer sitting on gene III (5′-CTCACTCATTAGGCA CCCAGGCTTTACAC-3′). Sequences containing in-frame ECRG4 fragments were translated and aligned with ECRG4 to identify the potential receptor binding domains.

PCR biopanning of phage internalization

The day before phage was added to PC-3 cells, the cells were seeded onto 100 mm tissue culture dishes at a concentration of 2.5×10^6 cells/plate. The following day, 1×10^{11} pfu/ml of the phage library was added drop-wise into the medium. After a 2-h incubation at 37 °C to enable receptor-mediated internalization, the cells were washed three times sequentially with PBS, 50 mM Glycine pH 2.8, and PBS to remove unbound, non-specifically bound and non-internalized particles. The medium was then refreshed and cells were incubated for another 8 h. Cells were then washed with PBS and DNA isolated using the Genomic DNA Buffer Set from Qiagen. Both cytosolic and nuclear DNA were precipitated, combined and used as PCR template. A pair of forward (pUCg3f: 5′-GGAAACAGCTATGACCATGATTACGCC-3′) and reverse (pUC250rev: 5′-TAGCGACAGAATCAA GTTTGCCTTTAGCG-3′) primers flanking the multiple cloning sites (MCS) in the pUC250 vector was employed to amplify internalized EGF fragments. Internalized ECRG4 fragments were amplified with the same forward primer (pUCg3f) but a reverse primer for the pUC198 plasmid (pUC198rev: 5′- CAGGTCAGACGATTGGCCTTGATA TTCAC-3′). The PCR reaction was performed for 35 cycles at 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 30 s and the PCR products were purified and cloned into either the TOPO pCR2.1 (EGF fragmentation) and the pDrive (ECRG4 fragmentation) cloning systems for sequencing.

Construction of ECRG4$^{31–70}$ phage

The in-frame internalized ECRG4 sequences recovered after biopanning were re-engineered with an extended linker sequence to optimize internalizing activity (Larocca et al. 2001, 2002). The ECRG4$^{31–70}$ fragment was amplified by PCR with a sense primer (5′- ATAAACCCATGGCGATGC TTCAAAAACGAG-3′) that contained a NcoI restriction

site and antisense primers that encoded a 5xGlycine repeat linker for flexibility, the original adapter sequence in the fragmentation library and a PstI restriction site (5′- TTC AGCTGCAGTTATCGGGCCGGCCGGCCGGCACCC CCTCCGCCACCGCCAAGGAATTCTTTGGCTT-3′). The PCR products were purified and cloned into the pUC198 vector as described above. Identity and authenticity of the construct was confirmed by DNA sequencing.

Immunocytochemistry of internalized phage particles

Cells were seeded in chamber slides at 1×10^4 cells/well the day before transduction and 1×10^{11} pfu/ml phage was incubated with the cells for 2 h. Cells were then washed sequentially with PBS three times, incubated with 50 mM Glycine pH 2.8 for 10 min at room temperature, and washed again with PBS. Cells were then permeabilized with methanol and fixed with fresh 2% paraformaldehyde/2% glucose in PBS for 20 min at room temperature. Cells were stained with anti-Fd of M13 bacteriophage polyclonal antibodies (Sigma B7786) and internalized phage particles detected by incubation with secondary Alexa 546 goat-anti-rabbit antibodies (Gonzalez et al. 2010). Imaging of cells was performed on an Olympus Fluoview 1000 laser scanning confocal microscope.

PCR and real-time quantitative PCR to detect phage internalization

To quantify internalization, the phage were diluted to 10^{11} pfu/ml in medium and added to PC-3 cells for 2 h at 37 °C. Phage particles bound non-specifically to cells were removed by washing six times with PBS, once with 50 mM glycine pH 2.8 with 0.5 M NaCl for 15 min at 37 °C and three more times with PBS. DNA was extracted by adding 5 mM Tris–Cl, pH 8.5 containing 5% NP-40 to cells and incubating at 70 °C for 15 min. The particle number of internalized phagemid in 1 ml of extract was determined by the standard curve method of qPCR using primers 5′- GGA AACAGCTATGACCATGATTACGCC-3′ and 5′-CAGGTC AGACGATTGGCCTTGATATTCAC-3′. Amplification was measured on a BioRad iQ5 Real-Time PCR instrument using iQ SYBR Green PCR master mix (BioRad, Hercules, CA) at 95 °C for 30 s, 60 °C for 30 s and 72 °C for 30 s.

Transient transfection

HEK293 cells were seeded the day before transfection at 6×10^5 in 60 mm dishes. The medium was refreshed with complete RPMI1640 free of antibiotics 2 h before transfection. A total of 5 µg of plasmids, 2.5 µg of pcDNA3.1-hECRG4 (Dang et al. 2012) or pcDNA3.1-hECRG4$\Delta^{31–70}$ and 2.5 µg of hu-TLR4 cDNA pDEST40 tagged with FLAG(Plasmid #42646, Addgene) were dissolved into 150 µl OPTI-MEM, which was mixed with 150 µl of OPTI-MEM premixed with 7.5 µl Lipofectamine 2000 and incubated for 5 min at room temperature (Life Technologies). The mixture was then incubated for another 20 min at room temperature, which was then added drop-wise into cells and cells were continued to incubate for 36 h for later experiments.

Measurement of TLR4-dependent activation of NFκb

Commercial human TLR4/NF-κB/SEAP reporter HEK-Blue™ cells were originally generated by co-transfection of HEK cells with human TLR4, MD-2 and CD14 co-receptor genes to express the innate immunity receptor complex and an inducible SEAP (secreted embryonic alkaline phosphatase) reporter gene to detect activation (InvivoGen, San Diego, CA). In these cells, the SEAP reporter gene was placed by the manufacturer under control of an IL-12 p40 minimal promoter that was fused to five NF-κB and AP-1-binding sites. Stimulation with a TLR4 ligand activates NF-κB and AP-1 to then induce the production of SEAP. The levels of SEAP were determined with a HEK-Blue™ Detection kit (InvivoGen, San Diego, CA) in cell culture selection medium that allows for real-time detection of SEAP. The parental cell line of HEK-Blue™ hTLR4 cells were HEK-Blue™ Null2 cells and SEAP activation measured as recommended by the manufacturer.

Immunoprecipitation

To confirm that the ECRG4$^{36–67}$ domain mediates an interaction between ECRG4 and the TLR4 complex, HEK293 cells were transiently co-transfected with plasmids over-expressing ECRG4 and TLR4-FLAG (Addgene) or ECRG4$^{\Delta 31–70}$ and TLR4-FLAG, and lysed with SDS lysis buffer (50 mM Tris–Cl, pH 7.0, 0.1% SDS, and 1 mM EDTA) 36 h after transfection. Lysate was centrifuged and equal amount of cell lysate immunoprecipitated with anti-ECRG4 monoclonal antibody overnight at 4 °C. The beads were the centrifuged and bound proteins recovered with SDS loading buffer and resolved on a 10% SDS-PAGE before transfer to a PVDF membrane and immunoblotting with anti-TLR4 (CAT#119079, Abcam) at a dilution of 1:5000 [44].

Results

Fragmentation and display of the EGF pre-pro-precursor open reading frame

The human EGF precursor gene consists of 24 exons spanning 110 kb in chromosome 4q25-q27. The EGF ORF

(Fig. 1a) encodes a large 1207 amino acid precursor that is, secreted, tethered at the cell surface and proteolytically processed to a small, 53-amino acid final product called EGF [47]. This mature EGF binds high affinity EGF receptors on the cell surface, internalizes into target cells and elicits many different biological effects [48, 49]. Because human EGF can transport phage into target cells through the EGF receptor when the ligand is displayed on phage [46, 50, 51], we used it to test the concept that phage display could identify the ligand encoded within the 3′ 200–400 base region of the EGF ORF.

To this end, the EGF precursor ORF was first amplified by PCR (Fig. 1b lane 7) and then fragmented by digesting the DNA with different concentrations of DNase I (Fig. 1b lane 2–6). Fragments of DNA ranging from 200 to 400 base pairs were then collected from the gel and cloned into a pUC250 phagemid vector that had been previously optimized for EGF display on phage and internalization of particles into EGF target cells [52]. The resulting phagemid libraries were estimated to have a minimal insert-less vector (< 10%) and a complexity of at least ~ 10^8 pfu. The phage library was prepared at a titer of 3.7×10^{13} pfu/ml titer of the phage as determined by ELISA [52].

Binding and internalization of pre-pro-EGF fragments into PC-3 cells

PC-3 cells are a prostate epithelial cancer cell line that express EGF receptors and efficiently internalize EGF-targeted phage [46, 53]. While there are numerous ways to recover internalizing phage, we used a PCR-based approach [54–56] to select for ligand internalization because this

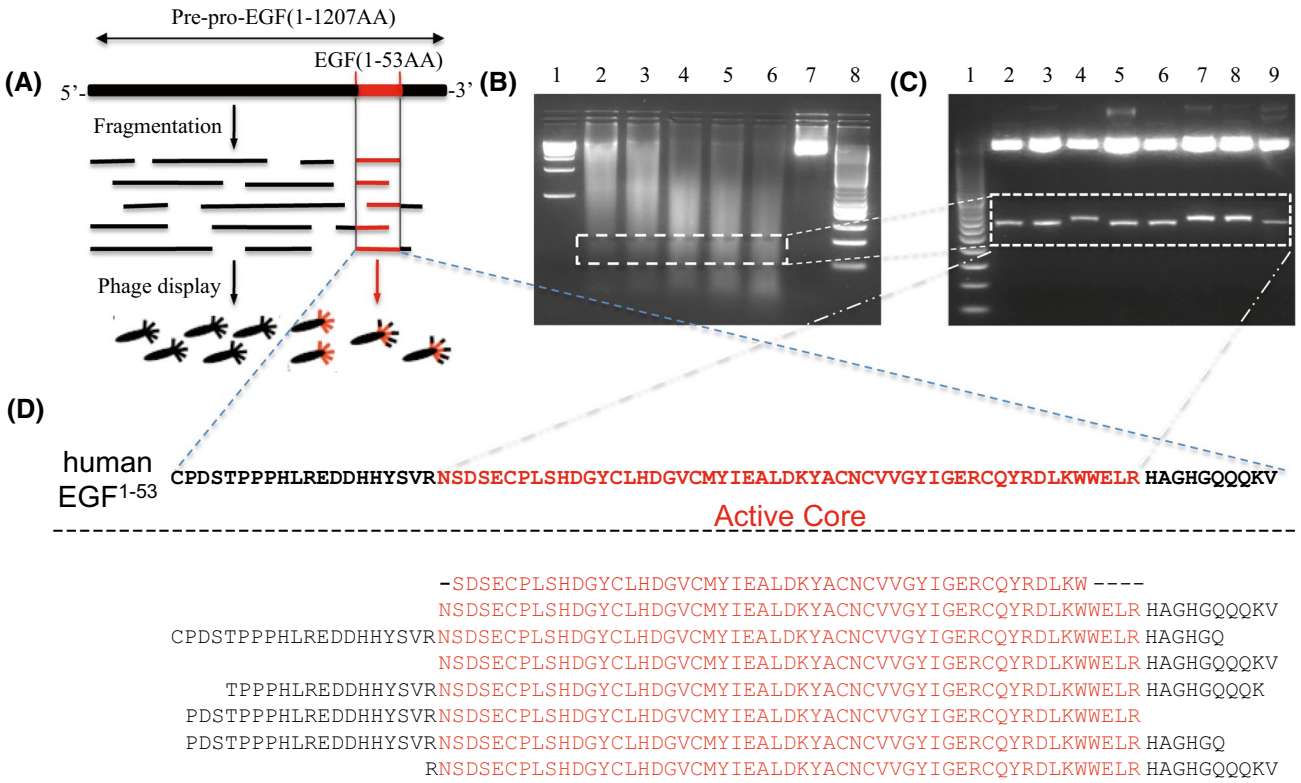

Fig. 1 Mining the human EGF precursor for the EGF ligand. **a** EGF Precursor processing: schematic representation of the human EGF pre-pro-precursor ORF (thick black solid line), mature EGF ligand (thick red line) and the randomly fragmented DNA fragments that were used to display on phage to create particles with EGR receptor targeting potential. **b** Optimization of human EGF ORF fragmentation by DNase I. PCR products were generated by digestion with either 1:8000 (lane 2), 4000 (lane 3), 2000 (lane 4), 1000 (lane 5), or 500 (lane 6) dilutions of DNase I (2 IU/ml) at 16 °C for 16 min and products compared to products of PCR with the intact EGF ORF (lane 7). By reference to the 1 kb (lane 1) and 100 bp (lane 8) DNA ladders, respectively. The 100–200 bp products (box) were selected for library construction. **c** Analyses of internalized phagemid: gel electrophoresis of 8 internalized oligonucleotide amplified fragments containing in-frame coding sequences of the active EGF ligand. The genomic phage DNA was recovered from transduced PC-3 cells, amplified by PCR, cloned into TOPO pCR2.1 and sequenced (lane 1 is 100 bp ladder). Eight plasmids (lanes 2–9) containing in-frame EGF-glll sequences were digested with EcoRI. **d** Alignment of human EGF ligand with recovered sequences. Sequencing the PCR products recovered from internalized phage DNA identified in-frame hEGF fragments that contained minimally active EGF sequences. The sequence of the authentic 53 amino acid human EGF peptide (in red and bold on the top) is aligned with the eight recovered EGF sequences (in red) and any short flanking EGF sequences (in black)

method has the advantage of selecting all internalized phage, including those that have lost bacterial infectivity as a result of exposure to intra-cellular proteases. PC-3 cells were incubated with 10^{11} pfu/ml phage from the EGF ORF-derived library and the internalized phage DNA recovered from cell extracts by PCR and cloned into bacteria. Eight random bacterial colonies were then picked after one round of screening, amplified and the phage DNA isolated for sequencing (Fig. 1c). Of these clones, sequencing identified in-frame sequences that contained both the EGF ORF and the Plll fusion partner that is derived from the filamentous phage genome. Translation and alignment of these sequences with the EGF precursor showed that they contained the core human EGF sequence (Fig. 1d).

Selection of internalizing ligands from the ECRG4 open reading frame

The human ECRG4 gene consists of 4 exons that span 12.5 kb on chromosome 2q12.2 (Fig. 2a). Transcription of this relatively small 447 bp ORF produces a 148 amino acid neuropeptide-like precursor protein (Fig. 2b) that bioinformatic analyses predict include a hydrophobic 30-amino acid leader-transmembrane peptide at the NH2 terminus for secretion processed by signal peptidase and six peptides ranging from 1 to 14 kDa [26] that are derived from the 17kDA ECRG4 and dependent on post-translational processing at protein convertase, furin and thrombin consensus cleavage sequences (Fig. 2b).

To explore whether the ORF ligand mining strategy could identify any of the eight ligand candidates in the orphan human ECRG4 gene, we displayed a library of ECRG4-fragmented oligonucleotides sized from 100 to 200 bp (Fig. 2c) in standard pUC198 phagemid particles. We then

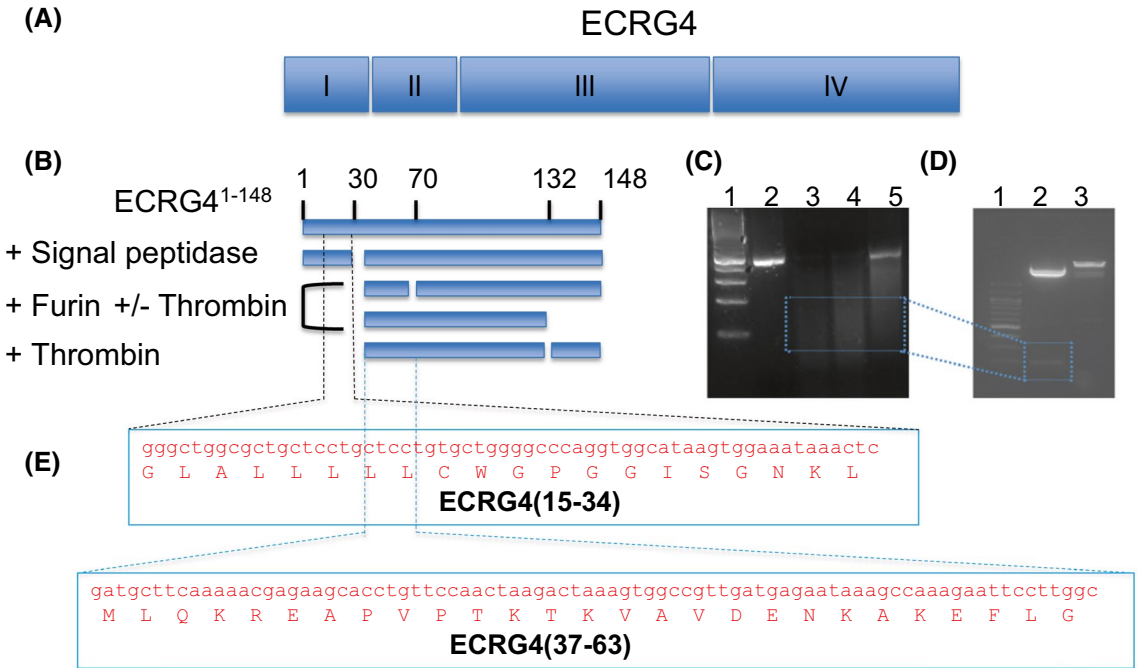

Fig. 2 Mining human ECRG4 Precursor for potential ligands with pUC250 phage. **a** ECRG4 precursor gene Located on chromosome 2 the gene is encoded by 4 exons. **b** Schematic representation of ECRG4 and its potential processing to smaller peptides. Intact human ECRG4 has 148 amino acids (including the initiation methionine encoded by ATG). It has a hydrophobic leader sequence for secretion and membrane tethering that can be cleaved by a signal peptidase and has predicted processing consensus sequences for combinations of furin and thrombin processing (black lines). **c** Fragmentation of human ECRG4 ORF by DNase I. The ORF was PCR amplified (one band, lane 2) purified and then digested at 16 °C for 15 min with DNase I at 1:50 dilution (lane 3), 1:100 (lane 4), 1:500 (lane 5) respectively. After analysis by gel electrophoresis, the 100–200 bp fraction highlighted on the gel was collected for insertion into phage. (lane 1 is 100 bp DNA ladder). **d** Restriction digestion of potential human ECRG4 ligands. Internalized ECRG4-targeted phage were analyzed by gel electrophoresis after digestion of plasmids with XhoI and BamHI. The 100 bp ladder (lane 1), ECRG4 fragments (lane 2) and lambda DNA/HindIII standards (lane 3) were resolved on 1.5% agarose gel and the 100–200 bp fraction highlighted on the gel used for plaquing and sequencing. **e** Sequences of 2 mined ECRG4 ligands from the randomly fragmented ECRG4 precursor ORF. Two classes of ECRG4 sequences were found to be in-frame with phage glll (red). They encode amino acids ECRG4$^{15–34}$ and ECRG4$^{37–63}$ of the original pre-pro-ECRG4 precursor sequence

selected for phage that internalize into PC3 cells, like shown with EGF (Fig. 1). As illustrated in Fig. 2d, 60 and 83 bp fragments were recovered from the PC-3 cell lysates and sequencing (Fig. 2e) identified ECRG4$^{15–34}$ and ECRG4$^{37–63}$, both in the amino terminal domain of the ECRG4 precursor. Predictably, the first (ECRG4$^{15–34}$) included the hydrophobic leader sequence that presumably internalized into PC3 cells because hydrophobic leader sequences can non-specifically penetrate cell membranes [57, 58]. The second peptide, however (ECRG4$^{37–63}$) infers the existence of a ligand binding sequence in ECRG4.

To confirm this sequence, we used a second independent phage library that was generated from fragmented sequences of < 100 bp to recover oligonucleotides encoding 15–20 amino acid peptides. After 3 rounds of biopanning (Table 1), 17 isolates carried in-frame ECRG4 sequences with 4 plaques (23.5%) containing the same ECRG4$^{37–63}$ core sequence selected from the previous biopanning strategy (Fig. 3a). In the 4th round biopanning, three of 4 in-frame isolates contained this core ECRG4 sequence. In this biopanning, a second domain in ECRG4$^{136–148}$ (Fig. 3b) selected in the 3rd round encodes a peptide derived from the C-terminus of ECRG4 called CT16 [24, 38, 59] and that was previously found to activate NFkB in target cells.

Internalization of ECRG4$^{37–63}$ into PC-3 Cells

To confirm that ECRG4$^{37–63}$ can internalize into PC-3 cells, the peptide was displayed with a flexible 5 glycine linker in pUC198 phagemid vector and PC-3 cells were then transduced with 10^{11} PFU/ml ECRG4$^{37–63}$ phage or with EGF-targeted phage. The internalized phage was

Table 1 Recovery of ligand targeted phage

	Total no. plaques sequenced	Plaques with ECRG4 sequences	Plaques with in-frame phage glll	Plaques with unrelated sequences	Plaques with no sequence inserted	% Plaques with ECRG4 (37–63)	% Plaques with ECRG4 (136–148)
Round 3	73	38	17	2	33	24 (4/17)	12 (2/17)
Round 4	24	11	4	0	9	75 (3/4)	0

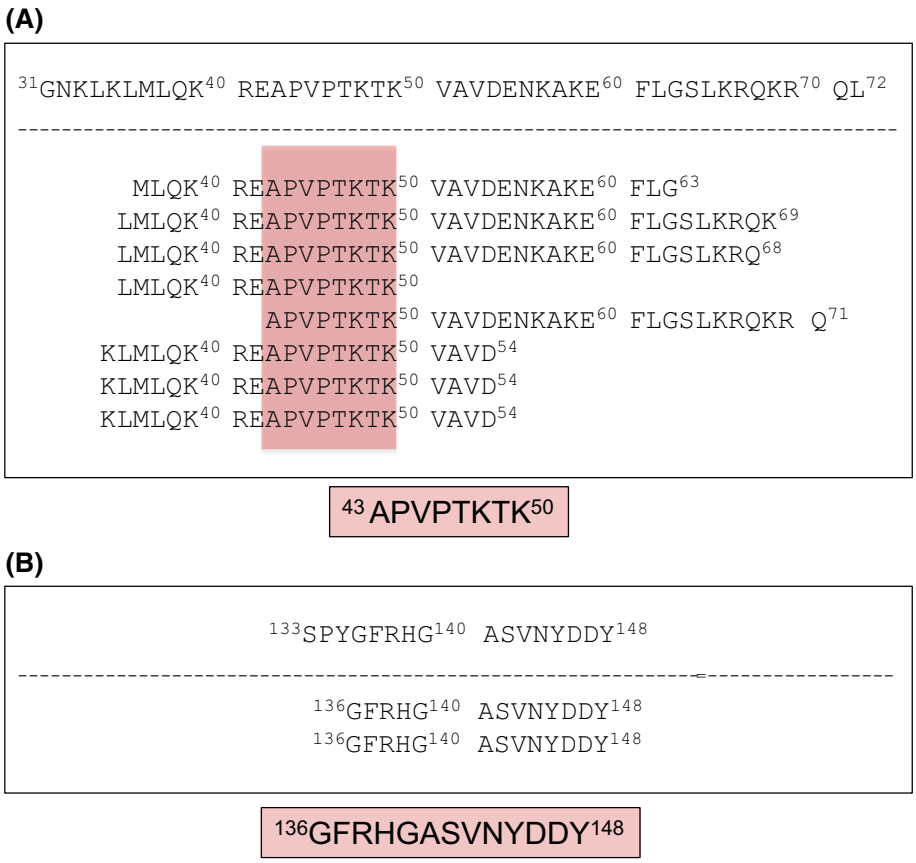

Fig. 3 Mining potential ligands from human c2orf40 with pUC198 phagemid. **a** Alignment of potential ECRG4 ligands from the amino terminus of ECRG4. Eight of the amino acid sequences align with the ECRG4$^{31–70}$ domain processed from the amino terminus of ECRG4. There is a common minimally active internalizing motif amongst all sequences that includes the eight amino acids of ECRG4$^{43–50}$ (shaded area). **b** Alignment of a potential carboxyl terminus ECRG4 ligand. The puc198 phage mined a 13 amino acid peptide sequence that aligned within the sequence of the ECRG4$^{133–148}$ domain predicted to be generated by thrombin processing

then detected by immunostaining PC3 cells with anti-phage antibody (Fig. 4). Little staining was detected in PC-3 cells incubated with untargeted phage prepared in the puc198 plasmid (Fig. 4a). In contrast, cells treated with the same concentration of ECRG4$^{37–63}$ phage in this plasmid showed intracellular staining (Fig. 4b). This ECRG4$^{37–63}$ phage staining was less than that observed with the EGF-targeted phage that was used as a positive control for internalization (Fig. 4c). We attribute this difference in staining intensity to the fact that the EGF-targeted phage were optimized for internalization in the puc250 vector (Burg et al. 2004). Indeed, when EGF is displayed in the pUC198 phage backbone, the internalization of both EGF- and ECRG4$^{31–70}$ -targeted phage is similar (Fig. 4d). Both ECRG4$^{31–70}$ and EGF-targeted phage were also found to target and internalize into normal dermal human fibroblasts (Fig. 4e), an ECRG4 target cell previously identified by Shaterian et al. [34].

Targeting is mediated by the TLR4-MD2-CD14 innate immunity receptor complex but does not activate NFκB

In previous studies, we used immunocytochemistry and co-immunoprecipitation to show that ECRG4 co-localizes with TLR4, MD2, and CD14 in human leukocytes [38] and later showed a direct interaction between ECRG4 and TLR4 [60]. To explore whether ECRG4$^{37–63}$ might mediate this interaction, 10^{11} PFU/ml ECRG4$^{37–63}$ phage were added to parental HEK293 cells or HEKblue™ cells that over-express TLR4, MD2, and CD14 components of the innate immunity complex. As shown in Fig. 5a–e, transduction of parental HEK293 cells did not show any measurable intracellular staining of untargeted, or EGF-, ECRG4$^{37–63}$, ECRG4$^{71–148}$ or ECRG4$^{133–148}$-targeted phage. In contrast, HEKblue™ cells transduced with ECRG4$^{37–63}$ and ECRG4$^{133–148}$-phage showed increased perinuclear staining compared to untargeted, EGF-, and ECRG4$^{71–148}$-targeted phage (Fig. 5f–j). This increased internalization of ECRG4$^{37–63}$ phage was confirmed by

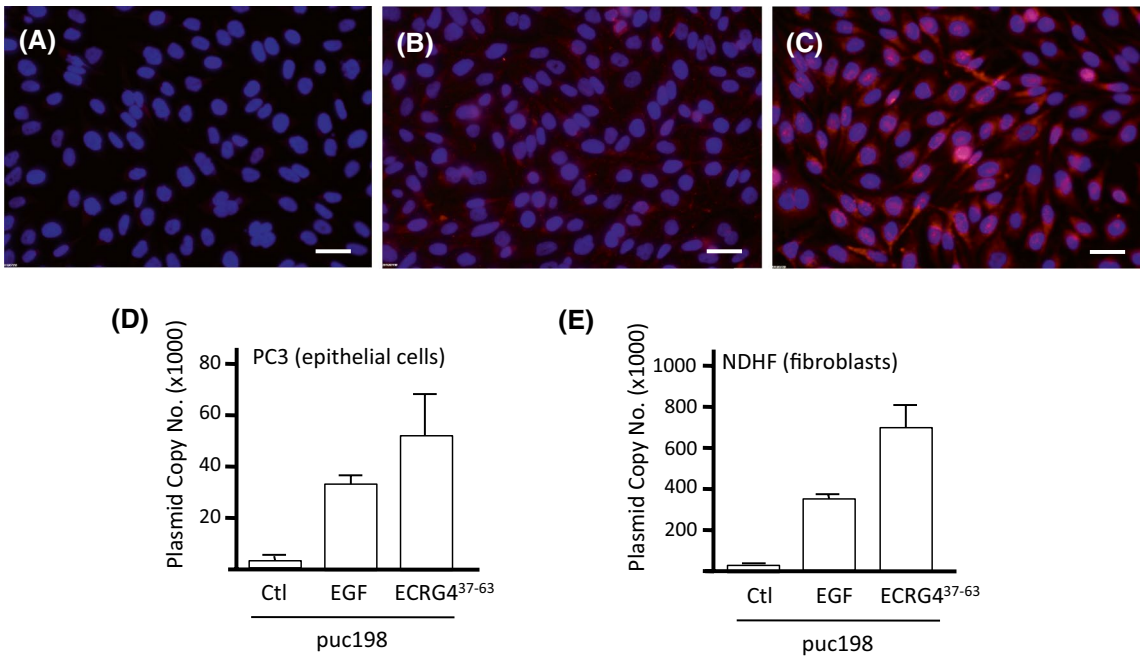

Fig. 4 Internalization of ECRG4$^{37–63}$-targeted phage into PC-3 cells. **a–c** Internalization of ECRG4 and EGF targeted phage. PC-3 cells were transduced with 10^{10} wild type untargeted phage grown from the pUC198 plasmid (panel **a**) or phage displaying ECRG4$^{37–63}$ in pUC198 (panel **b**) and compared to a human EGF displayed grown from the optimized pUC250 plasmid (panel **c**). Cells were washed with acid, formalin fixed and immunostained with antibodies to M13 phage coat protein to show the internalized phage particles (red) and its intracellular localization was compared with DAPI-stained nuclei (blue) by merging photomicrographs acquired by Olympus Fluoview confocal microscopy. **d** Targeting PC3 cells with phagemid and quantification by PCR: after incubating PC-3 cells with 10^{10} untargeted puc198 phage (Ctl) or EGF-targeted and ECRG4$^{37–63}$-targeted phage, cells were lysed and phage DNA amplified by real-time PCR. Plasmid copy number was measured using a standard curve generated by PCR of purified puc198 plasmid. **e** ECRG4$^{37–63}$ can target phagemid to fibroblasts: normal human fibroblasts (NHDF) were incubated for 2 h at 37C with either 10^{10} untargeted phage (Ctl), EGF-targeted or ECRG4$^{37–63}$-targeted phage grown from the puc198 plasmid. The cells were lysed and phage DNA amplified by real-time PCR to quantify phage DNA internalization

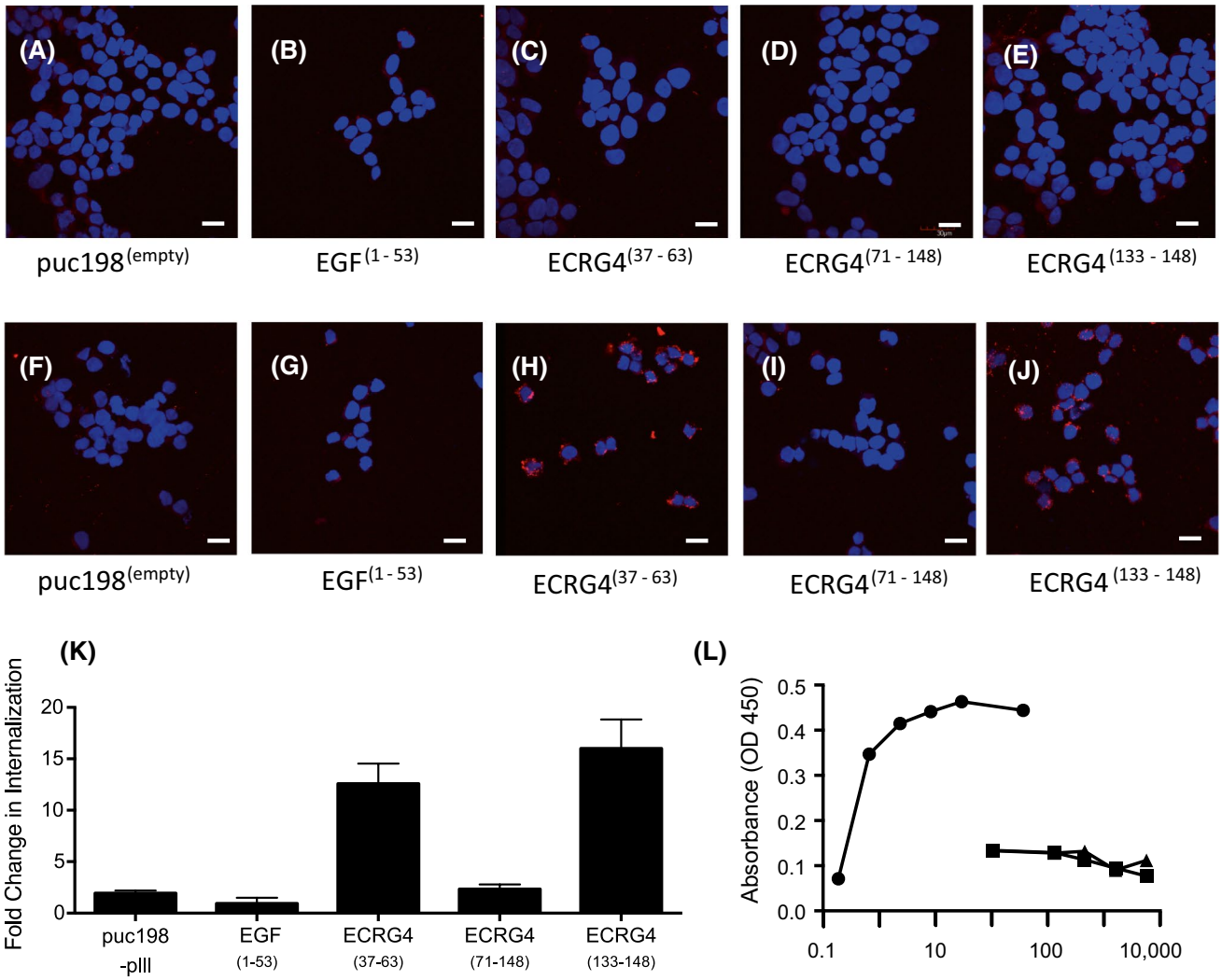

Fig. 5 Internalization of the ECRG4$^{37–63}$-targeted phage into reporter HEK293 cells engineered to express TLR4-MD2-CD14 innate immunity complex. Panel **a**–**e** Targeting control HEK293 cells Wild type HEK cells were incubated with 10^{11} PFU/ml of wild type pUC198 (empty), EGF$^{1–53}$-targeted pUC198, ECRG4$^{37–63}$-targeted pUC198, ECRG4$^{71–148}$-targeted pUC198 or ECRG4$^{133–148}$-targeted pUC198 phage, respectively. Cells were fixed and phage internalization was visualized by immunofluorescence using an anti-phage antibody. Nuclei (blue) stained with DAPI. Panel **f**–**j** Targeting TLR4 expressing HEK cells. HEKblue™ cells with the innate immunity receptor complex were incubated with 10^{11} PFU/ml of wild-type pUC198 (empty), EGF$^{1–53}$-targeted pUC198, ECRG4$^{37–63}$-targeted pUC198, Ecrg4$^{71–148}$-targeted pUC198 or ECRG4$^{133–148}$-targeted pUC198 phage, respectively. Cells were fixed and phage internalization was visualized by immunofluorescence using an anti-phage antibody. The internalization into parental HEK293 cells was minimal for all phages. However, strong perinuclear staining (red) was observed in HEKblue™ cells incubated with ECRG4$^{37–63}$-targeted pUC198 (panel **h**) and ECRG4$^{133–148}$-targeted pUC198 phage (panel **j**). Nuclei (blue) stained with DAPI. Panel **k** Quantification of phage internalization into HEKblue cells by real-time PCR. HEKblue™ cells were incubated with 10^{11} PFU/ml of the pUC198 phages indicated and DNA of internalized phage DNA was quantified by real-time PCR. The internalization was normalized to that of parental HEK293 cells. Data were presented as mean ± standard deviation, and experiments were repeated at least three times in triplicate. Panel **l**: Activation of NFkB in HEKblue™ cells. HEKblue™ cells were incubated with the indicated concentrations (ng/ml) of lipopolysaccharide (circles), recombinant ECRG4$^{31–70}$ (squares) or ECRG4$^{133–148}$ (triangles) and the production of SEAP measured by optical density

real-time quantitative PCR that showing the internalization of phage into HEKblue™ cells being 13-fold higher than the other phage tested (Fig. 5k). Neither peptide activated NKκb although lipopolysaccharide was clearly active at very low concentrations (Fig. 5k).

A sequence within the ECRG4$^{31–70}$ domain interacts with the TLR4-MD2-CD14 complex

HEK293 cells were co-transfected with a plasmid encoding a TLR4-FLAG fusion protein and either a plasmid encoding either full-length ECRG4$^{1–148}$ or the ECRG4$^{\Delta 30–70}$ mutant

that was engineered to lack the 31–70 amino acid sequence. Because intact ECRG4$^{1–148}$ is known to interact with the innate immunity receptor complex leukocytes [38, 60], the anti-ECRG4 immuno-precipitated TLR4 and the anti-FLAG antibodies that immuno-precipitate TLR4 also immuno-precipitate ECRG4 (Fig. 6). In contrast, the in-frame deletion of amino acids ECRG4$^{31–70}$ from the primary structure of ECRG4 significantly decreased the interactions between ECRG4 and TLR4. These results indicate that the interaction between ECRG4 and the innate immunity complex is mediated in large part by the ECRG4$^{31–70}$ amino terminus sequence. The remaining signal is likely due to the interaction with ECRG4$^{133–148}$ with the innate immunity receptor published previously [38, 60].

Discussion

Since the discovery of phage display in the mid 1980s [61], it has been used for numerous applications including the mapping of functional domains of proteins [62–64]. Here, the technique has been adapted to mine ligand-like domains in an orphan neuropeptide-like pre-pro-precursor open reading frame (ORF). The use of a fragmented cDNA libraries derived from pre-defined ORFs has significant advantages in domain identification, not the least of which is their lower complexity and the increased representation of putative ligands during biopanning [65, 66]. Moreover, the combinatorial complexity of the library can be decreased by pre-selecting the display of in-frame cDNA fragments during library construction [56]. Out-of-frame false positive ligand sequences are minimized and eliminated by informatic analyses because in-frame cDNA fragments are required to display both potential ligand and phage plll protein. As shown here, all in-frame sequences identified after mining the EGF precursor ORF (Fig. 1) contained an EGF core sequence and could be immediately distinguished from false positives in only one round of biopanning. With up to four rounds of biopanning we identified also three distinct internalizing domains in ECRG4 that include (1) its hydrophobic signal peptide, (2) an amino terminus within ECRG4$^{31–70}$ domain and (3) a carboxyl-terminus peptide within the ECRG4$^{133–148}$ domain. Because the latter has been previously been shown by us [24, 38] to activate NFκb in target cells and the Kondo laboratories [59] have shown it to enter cells via the scavenger receptor, we sought to identify the sequence that accounts for interactions between ECRG4 and the TLR4 innate receptor complex [60]. This interaction appears to be mediated by a peptide sequence in the ECRG4$^{31–70}$ amino terminus domain of ECRG4 (Fig. 2b).

The possibility that an ECRG4 precursor encodes ligand-like peptides and collectively encoded by the c2orf40 gene was initially predicted by Markov modelling that led Mirabeau to suggest the existence of a new hormone-like peptide called augurin [26]. The very existence and identity of augurin, however, soon came into question when systematic analyses of conditioned media from cells in vitro [43, 44] and later in vivo [38], as well as mutational analysis of ECRG4 [44] showed that there was cell-specific and complex proteolytic processing at combinations of thrombin, furin and protease convertase sites that generate several ECRG4-derived peptides (reviewed in [39]). One such peptide, CT16, is located at the C-terminus of ECRG4, generated by C-terminal processing of ECRG4 by thrombin for release and activates NFkB to promote leukocyte migration in vitro and in vivo [24, 38]. This same domain was

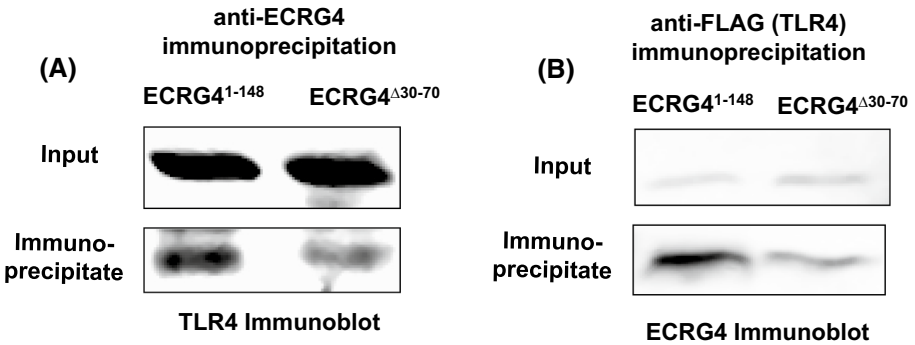

Fig. 6 ECRG4$^{37–63}$ interacts with TLR4 in the TLR4-CD14-MD2 innate immune complex. **a** Anti ECRG4 immunoprecipitation of TLR4: HEK293 cells were transiently transfected for 36 h with either ECRG4 or the ECRG4$^{\Delta 31–70}$ lacking the active domain, along with a TLR4-FLAG plasmid. Cells were then lysed and ECRG4 immunoprecipitated with anti-ECRG4 for immunoblotting with an anti-TLR4 antibody. Input was loaded 1/30 of the amount used in immunoprecipitation and used to demonstrate equivalent starting material with each sample. **b** Anti-TLR4 immunoprecipitation of ECRG4: HEK293 cells were transiently transfected for 36 h with either ECRG4 or ECRG4$^{\Delta 31–70}$ and TLR4-FLAG. Cells were lysed and protein immunoprecipitated with anti-FLAG antibodies that precipitate TLR4 which was then immunoblotted with anti-ECRG4 antibodies. Input was loaded 1/30 of the amount used in immunoprecipitation and used to demonstrate equivalent starting material with each sample

identified here by biopanning PC-3 cells (Fig. 3), a result that confirms recently reported findings of Moriguchiu [59] using retrovirus-mediated expression cloning. In contrast, the ECRG4$^{31–70}$ domain identified here contains a core 36 amino acid sequence (LMLQKREAPVPTKTKVAVD) that is sufficient for internalization into cells. It is likely the same amino terminus processed peptide detected, but not identified, by Ozawa et al. [43] in the conditioned media of cells transfected with ECRG4 cDNA. Interestingly, we did not detect peptides in the ECRG4$^{71–148}$ domain that Moriguchi [59] identified as a ligand for the LOX-1 scavenger receptor. Instead, we ascribe the ability of ECRG4 to interact with TLR4 with sequences in the ECRG4$^{31–70}$ domain. Assuming that LOX-1 is underrepresented in PC3 cells, the findings underscore the importance of the target cell in selecting a screening strategy for orphan ligands.

It is interesting to speculate that the in vitro and in vivo regulation of the ECRG4 gene expression in experimental models of cancer, tumor progression and metastases, CNS responsiveness to injury, cell senescence, mucosal proliferation and neuroprogenitor cell differentiation (reviewed in [39, 42]) might regulate the activities of one or more of the peptide products of ECRG4 gene expression described here. ECRG4$^{71–148}$ is reported to stimulate CRF release [29] while both ECRG4$^{71–132}$ and ECRG4$^{133–148}$ interact with the LOX1 scavenger receptor [40, 59]. Here, we posit that a peptide sequence represented by APVPTKTK in the ECRG4$^{31–70}$ domain accounts for the reported interaction between ECRG4 and the innate immunity receptor complex [60]. That being said, the inability of these peptides to activate NFκB in the same NFκB reporter cell system that binds and internalizes ECRG4 may point to involvement of the non-canonical pathway of NFκB activation after innate immunity receptor complex activation or the existence of a heretofore unidentified pathway that mediates the effects of ECRG4 on cell migration, growth and differentiated function.

Acknowledgements Research originally supported by the National Institutes of Health P20-GM078421 (AB/BPE), EY018479 (AB), HL73396 (BPE), the American Recovery Act (ARRA), and subsequently from GM121530 (TC), UC San Diego Department of Surgery Reinvestment Fund (RC) and completed with funding from grants 81-770-336 and 30-870-903 from the National Natural Science Foundation of China (XD). The authors declare no conflicts of interest and are indebted to Emelie Amburn's expert laboratory support.

References

1. Lander ES (2011) Initial impact of the sequencing of the human genome. Nature 470(7333):187–197. https://doi.org/10.1038/nature09792
2. McManus CJ, Graveley BR (2011) RNA structure and the mechanisms of alternative splicing. Curr Opin Genet Dev 21(4):373–379. https://doi.org/10.1016/j.gde.2011.04.001
3. Poulos MG, Batra R, Charizanis K, Swanson MS (2011) Developments in RNA splicing and disease. Cold Spring Harb Perspect Biol 3(1):a000778. https://doi.org/10.1101/cshperspect.a000778
4. Castro MG, Morrison E (1997) Post-translational processing of proopiomelanocortin in the pituitary and in the brain. Crit Rev Neurobiol 11(1):35–57
5. Hanada K, Yang JC (2005) Novel biochemistry: post-translational protein splicing and other lessons from the school of antigen processing. J Mol Med (Berl) 83(6):420–428. https://doi.org/10.1007/s00109-005-0652-6
6. Falth M, Skold K, Svensson M, Nilsson A, Fenyo D, Andren PE (2007) Neuropeptidomics strategies for specific and sensitive identification of endogenous peptides. Mol Cell Proteom 6(7):1188–1197. https://doi.org/10.1074/mcp.M700016-MCP200
7. Svensson M, Skold K, Nilsson A, Falth M, Nydahl K, Svenningsson P, Andren PE (2007) Neuropeptidomics: MS applied to the discovery of novel peptides from the brain. Anal Chem 79(1):15–16
8. Bystroff C, Krogh A (2008) Hidden Markov Models for prediction of protein features. Methods Mol Biol 413:173–198
9. Schuster-Bockler B, Bateman A (2007) An introduction to hidden Markov models. Curr Protoc Bioinform Append 3:3. https://doi.org/10.1002/0471250953.bia03as18
10. Wu J, Xie J (2010) Hidden Markov model and its applications in motif findings. Methods Mol Biol 620:405–416. https://doi.org/10.1007/978-1-60761-580-4_13
11. Hathout Y (2007) Approaches to the study of the cell secretome. Expert Rev Proteom 4(2):239–248. https://doi.org/10.1586/14789450.4.2.239
12. Mustafa SA, Hoheisel JD, Alhamdani MS (2011) Secretome profiling with antibody microarrays. Mol BioSyst 7(6):1795–1801. https://doi.org/10.1039/c1mb05071k
13. Bi MX, Han WD, Lu SX (2001) Using Lab On-line to Clone and Identify the Esophageal Cancer Related Gene 4. Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Biochem Biophys Bull Shanghai) 33(3):257–261
14. Yue CM, Deng DJ, Bi MX, Guo LP, Lu SH (2003) Expression of ECRG4, a novel esophageal cancer-related gene, downregulated by CpG island hypermethylation in human esophageal squamous cell carcinoma. World J Gastroenterol 9(6):1174–1178
15. Cai Z, Liang P, Xuan J, Wan J, Guo H (2016) ECRG4 as a novel tumor suppressor gene inhibits colorectal cancer cell growth in vitro and in vivo. Tumour Biol 37(7):9111–9120. https://doi.org/10.1007/s13277-015-4775-2
16. Chen J, Liu C, Yin L, Zhang W (2015) The tumor-promoting function of ECRG4 in papillary thyroid carcinoma and its related mechanism. Tumour Biol 36(2):1081–1089. https://doi.org/10.1007/s13277-014-2731-1
17. Chen JY, Wu X, Hong CQ, Chen J, Wei XL, Zhou L, Zhang HX, Huang YT, Peng L (2017) Downregulated ECRG4 is correlated with lymph node metastasis and predicts poor outcome for nasopharyngeal carcinoma patients. Clin Transl Oncol 19(1):84–90. https://doi.org/10.1007/s12094-016-1507-z
18. Gotze S, Feldhaus V, Traska T, Wolter M, Reifenberger G, Tannapfel A, Kuhnen C, Martin D, Muller O, Sievers S (2009) ECRG4 is a candidate tumor suppressor gene frequently hypermethylated in colorectal carcinoma and glioma. BMC Cancer 9:447. https://doi.org/10.1186/1471-2407-9-447
19. Matsuzaki J, Torigoe T, Hirohashi Y, Tamura Y, Asanuma H, Nakazawa E, Saka E, Yasuda K, Takahashi S, Sato N (2013) Expression of ECRG4 is associated with lower proliferative potential of esophageal cancer cells. Pathol Int 63(8):391–397. https://doi.org/10.1111/pin.12079
20. Mori Y, Ishiguro H, Kuwabara Y, Kimura M, Mitsui A, Kurehara H, Mori R, Tomoda K, Ogawa R, Katada T, Harata K, Fujii Y (2007) Expression of ECRG4 is an independent prognostic factor

for poor survival in patients with esophageal squamous cell carcinoma. Oncol Rep 18(4):981–985
21. Jia J, Dai S, Sun X, Sang Y, Xu Z, Zhang J, Cui X, Song J, Guo X (2015) A preliminary study of the effect of ECRG4 overexpression on the proliferation and apoptosis of human laryngeal cancer cells and the underlying mechanisms. Mol Med Rep 12(4):5058–5064. https://doi.org/10.3892/mmr.2015.4059
22. Jiang CP, Wu BH, Wang BQ, Fu MY, Yang M, Zhou Y, Liu F (2013) Overexpression of ECRG4 enhances chemosensitivity to 5-fluorouracil in the human gastric cancer SGC-7901 cell line. Tumour Biol 34(4):2269–2273. https://doi.org/10.1007/s13277-013-0768-1
23. Deng P, Chang XJ, Gao ZM, Xu XY, Sun AQ, Li K, Dai DQ (2018) Downregulation and DNA methylation of ECRG4 in gastric cancer. Onco Targets Ther 11:4019–4028. https://doi.org/10.2147/OTT.S161200
24. Lee J, Dang X, Borboa A, Coimbra R, Baird A, Eliceiri BP (2015) Thrombin-processed Ecrg4 recruits myeloid cells and induces antitumorigenic inflammation. Neuro Oncol 17(5):685–696. https://doi.org/10.1093/neuonc/nou302
25. Li L, Zhang C, Li X, Lu S, Zhou Y (2010) The candidate tumor suppressor gene ECRG4 inhibits cancer cells migration and invasion in esophageal carcinoma. J Exp Clin Cancer Res 29:133. https://doi.org/10.1186/1756-9966-29-133
26. Mirabeau O, Perlas E, Severini C, Audero E, Gascuel O, Possenti R, Birney E, Rosenthal N, Gross C (2007) Identification of novel peptide hormones in the human proteome by hidden Markov model screening. Genome Res 17(3):320–327. https://doi.org/10.1101/gr.5755407
27. Gonzalez AM, Podvin S, Lin SY, Miller MC, Botfield H, Leadbeater WE, Roberton A, Dang X, Knowling SE, Cardenas-Galindo E, Donahue JE, Stopa EG, Johanson CE, Coimbra R, Eliceiri BP, Baird A (2011) Ecrg4 expression and its product augurin in the choroid plexus: impact on fetal brain development, cerebrospinal fluid homeostasis and neuroprogenitor cell response to CNS injury. Fluids Barriers CNS 8(1):6. https://doi.org/10.1186/2045-8118-8-6
28. Podvin S, Gonzalez AM, Miller MC, Dang X, Botfield H, Donahue JE, Kurabi A, Boissaud-Cooke M, Rossi R, Leadbeater WE, Johanson CE, Coimbra R, Stopa EG, Eliceiri BP, Baird A (2011) Esophageal cancer related gene-4 is a choroid plexus-derived injury response gene: evidence for a biphasic response in early and late brain injury. PLoS One 6(9):e24609. https://doi.org/10.1371/journal.pone.0024609
29. Tadross JA, Patterson M, Suzuki K, Beale KE, Boughton CK, Smith KL, Moore S, Ghatei MA, Bloom SR (2010) Augurin stimulates the hypothalamo-pituitary-adrenal axis via the release of corticotrophin-releasing factor in rats. Br J Pharmacol 159(8):1663–1671. https://doi.org/10.1111/j.1476-5381.2010.00655.x
30. Matsuzaki J, Torigoe T, Hirohashi Y, Kamiguchi K, Tamura Y, Tsukahara T, Kubo T, Takahashi A, Nakazawa E, Saka E, Yasuda K, Takahashi S, Sato N (2012) ECRG4 is a negative regulator of caspase-8-mediated apoptosis in human T-leukemia cells. Carcinogenesis 33(5):996–1003. https://doi.org/10.1093/carcin/bgs118
31. Huh YH, Ryu JH, Shin S, Lee DU, Yang S, Oh KS, Chun CH, Choi JK, Song WK, Chun JS (2009) Esophageal cancer related gene 4 (ECRG4) is a marker of articular chondrocyte differentiation and cartilage destruction. Gene 448(1):7–15. https://doi.org/10.1016/j.gene.2009.08.015
32. Kujuro Y, Suzuki N, Kondo T (2010) Esophageal cancer-related gene 4 is a secreted inducer of cell senescence expressed by aged CNS precursor cells. Proc Natl Acad Sci USA 107(18):8259–8264. https://doi.org/10.1073/pnas.0911446107
33. Kao S, Shaterian A, Cauvi DM, Dang X, Chun HB, De Maio A, Costantini TW, Coimbra R, Eliceiri BP, Baird A (2015) Pulmonary preconditioning, injury, and inflammation modulate expression of the candidate tumor suppressor gene ECRG4 in lung. Exp Lung Res 41(3):162–172. https://doi.org/10.3109/01902148.2014.983282
34. Shaterian A, Kao S, Chen L, DiPietro LA, Coimbra R, Eliceiri BP, Baird A (2013) The candidate tumor suppressor gene Ecrg4 as a wound terminating factor in cutaneous injury. Arch Dermatol Res 305(2):141–149. https://doi.org/10.1007/s00403-012-1276-7
35. Kurabi A, Pak K, Dang X, Coimbra R, Eliceiri BP, Ryan AF, Baird A (2013) Ecrg4 attenuates the inflammatory proliferative response of mucosal epithelial cells to infection. PLoS One 8(4):e61394. https://doi.org/10.1371/journal.pone.0061394
36. Nakatani Y, Kiyonari H, Kondo T (2019) Ecrg4 deficiency extends the replicative capacity of neural stem cells in a Foxg1-dependent manner. Development. https://doi.org/10.1242/dev.168120
37. Nishikawa M, Drmanac RT, Lobal I, Tang Y, Lee J, Stache-Crain B (2008) Polypeptide having an activity to support proliferation or survival of hematopoietic stem cell and hematopoietic progenitor cell, and DNA coding for the same. United States Patent Trade Office USA, Nuvelo Inc, San Carlos California
38. Baird A, Coimbra R, Dang X, Lopez N, Lee J, Krzyzaniak M, Winfield R, Potenza B, Eliceiri BP (2012) Cell surface localization and release of the candidate tumor suppressor Ecrg4 from polymorphonuclear cells and monocytes activate macrophages. J Leukoc Biol 91(5):773–781. https://doi.org/10.1189/jlb.1011503
39. Baird A, Lee J, Podvin S, Kurabi A, Dang X, Coimbra R, Costantini T, Bansal V, Eliceiri BP (2014) Esophageal cancer-related gene 4 at the interface of injury, inflammation, infection, and malignancy. Gastrointest Cancer 4:131–142. https://doi.org/10.2147/GICTT.S49085
40. Moriguchi T, Kaneumi S, Takeda S, Enomoto K, Mishra SK, Miki T, Koshimizu U, Kitamura H, Kondo T (2016) Ecrg4 contributes to the anti-glioma immunosurveillance through type-I interferon signaling. Oncoimmunology 5(12):e1242547. https://doi.org/10.1080/2162402X.2016.1242547
41. Porzionato A, Rucinski M, Macchi V, Sarasin G, Malendowicz LK, De Caro R (2015) ECRG4 expression in normal rat tissues: expression study and literature review. Eur J Histochem 59(2):2458. https://doi.org/10.4081/ejh.2015.2458
42. Qin X, Zhang P (2018) ECRG4: a new potential target in precision medicine. Front Med. https://doi.org/10.1007/s11684-018-0637-9
43. Ozawa A, Lick AN, Lindberg I (2011) Processing of proaugurin is required to suppress proliferation of tumor cell lines. Mol Endocrinol 25(5):776–784. https://doi.org/10.1210/me.2010-0389
44. Dang X, Podvin S, Coimbra R, Eliceiri B, Baird A (2012) Cell-specific processing and release of the hormone-like precursor and candidate tumor suppressor gene product, Ecrg4. Cell Tissue Res 348(3):505–514. https://doi.org/10.1007/s00441-012-1396-6
45. Dang X, Zeng X, Coimbra R, Eliceiri BP, Baird A (2017) Counter regulation of ECRG4 gene expression by hypermethylation-dependent inhibition and the Sp1 transcription factor-dependent stimulation of the c2orf40 promoter. Gene 636:103–111. https://doi.org/10.1016/j.gene.2017.08.041
46. Larocca D, Kassner PD, Witte A, Ladner RC, Pierce GF, Baird A (1999) Gene transfer to mammalian cells using genetically targeted filamentous bacteriophage. FASEB J 13(6):727–734
47. Rousselet E, Benjannet S, Marcinkiewicz E, Asselin MC, Lazure C, Seidah NG (2011) Proprotein convertase PC7 enhances the activation of the EGF receptor pathway through processing of the EGF precursor. J Biol Chem 286(11):9185–9195. https://doi.org/10.1074/jbc.M110.189936
48. Rall LB, Scott J, Bell GI, Crawford RJ, Penschow JD, Niall HD, Coghlan JP (1985) Mouse prepro-epidermal growth factor synthesis by the kidney and other tissues. Nature 313(5999):228–231

49. Xian CJ, Zhou XF (2004) EGF family of growth factors: essential roles and functional redundancy in the nerve system. Front Biosci 9:85–92
50. Larocca D, Baird A (2001) Receptor-mediated gene transfer by phage-display vectors: applications in functional genomics and gene therapy. Drug Discov Today 6(15):793–801
51. Larocca D, Witte A, Johnson W, Pierce GF, Baird A (1998) Targeting bacteriophage to mammalian cell surface receptors for gene delivery. Hum Gene Ther 9(16):2393–2399. https://doi.org/10.1089/hum.1998.9.16-2393
52. Larocca D, Jensen-Pergakes K, Burg MA, Baird A (2001) Receptor-targeted gene delivery using multivalent phagemid particles. Mol Ther 3(4):476–484. https://doi.org/10.1006/mthe.2001.0284
53. Larocca D, Burg MA, Jensen-Pergakes K, Ravey EP, Gonzalez AM, Baird A (2002) Evolving phage vectors for cell targeted gene delivery. Curr Pharm Biotechnol 3(1):45–57
54. Burg M, Ravey EP, Gonzales M, Amburn E, Faix PH, Baird A, Larocca D (2004) Selection of internalizing ligand-display phage using rolling circle amplification for phage recovery. DNA Cell Biol 23(7):457–462. https://doi.org/10.1089/1044549041474760
55. Burg MA, Jensen-Pergakes K, Gonzalez AM, Ravey P, Baird A, Larocca D (2002) Enhanced phagemid particle gene transfer in camptothecin-treated carcinoma cells. Cancer Res 62(4):977–981
56. Faix PH, Burg MA, Gonzales M, Ravey EP, Baird A, Larocca D (2004) Phage display of cDNA libraries: enrichment of cDNA expression using open reading frame selection. Biotechniques 36(6):1018–1022
57. Joliot A (2005) Transduction peptides within naturally occurring proteins. Sci STKE. https://doi.org/10.1126/stke.3132005pe54
58. Peitz M, Pfannkuche K, Rajewsky K, Edenhofer F (2002) Ability of the hydrophobic FGF and basic TAT peptides to promote cellular uptake of recombinant Cre recombinase: a tool for efficient genetic engineering of mammalian genomes. Proc Natl Acad Sci USA 99(7):4489–4494. https://doi.org/10.1073/pnas.032068699
59. Moriguchi T, Takeda S, Iwashita S, Enomoto K, Sawamura T, Koshimizu U, Kondo T (2018) Ecrg4 peptide is the ligand of multiple scavenger receptors. Sci Rep 8(1):4048. https://doi.org/10.1038/s41598-018-22440-4
60. Podvin S, Dang X, Meads M, Kurabi A, Costantini T, Eliceiri BP, Baird A, Coimbra R (2015) Esophageal cancer-related gene-4 (ECRG4) interactions with the innate immunity receptor complex. Inflamm Res 64(2):107–118. https://doi.org/10.1007/s00011-014-0789-2
61. Smith GP (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science 228(4705):1315–1317
62. Jespers LS, Messens JH, De Keyser A, Eeckhout D, Van den Brande I, Gansemans YG, Lauwereys MJ, Vlasuk GP, Stanssens PE (1995) Surface expression and ligand-based selection of cDNAs fused to filamentous phage gene VI. Biotechnology (NY) 13(4):378–382
63. Smith GP, Petrenko VA (1997) Phage display. Chem Rev 97(2):391–410
64. Thompson J, Pope T, Tung JS, Chan C, Hollis G, Mark G, Johnson KS (1996) Affinity maturation of a high-affinity human monoclonal antibody against the third hypervariable loop of human immunodeficiency virus: use of phage display to improve affinity and broaden strain reactivity. J Mol Biol 256(1):77–88. https://doi.org/10.1006/jmbi.1996.0069
65. Kramer RA, Cox F, van der Horst M, van der Oudenrijn S, Res PC, Bia J, Logtenberg T, de Kruif J (2003) A novel helper phage that improves phage display selection efficiency by preventing the amplification of phages without recombinant protein. Nucleic Acids Res 31(11):e59
66. Smothers JF, Henikoff S, Carter P (2002) Tech.Sight. Phage display. Affinity selection from biological libraries. Science 298(5593):621–622. https://doi.org/10.1126/science.298.5593.621