DEGREE PROJECT IN BIOTECHNOLOGY,
SECOND CYCLE, 30 CREDITS
*STOCKHOLM, SWEDEN 2021*

# Characterization of novel bispecific ADAPTs selected for cancer-related targets

**BLENDA HEDIN**

**KTH ROYAL INSTITUTE OF TECHNOLOGY**
**SCHOOL OF ENGINEERING SCIENCES IN CHEMISTRY, BIOTECHNOLOGY AND HEALTH**

# ABSTRACT

Cancer is still one of the most common causes of death world-wide and in parallel there is a need to update the repertoire of therapies that withstand resistance of recurrent cancers. Since the introduction of antibody therapies as anti-cancer pharmaceuticals, recognized as immunotherapy in health care, it has been an increasing field in cancer therapy, as a more targeted treatment compared to chemotherapy. Despite the great success, immunotherapy rely on parenteral administration, partly due to poor tissue penetration. If the treatment is administered intravenously, specialized personnel is required, in addition to that it can be inconvenient for the patient. Also, pharmaceuticals based on antibodies often require costly production steps which yields a high-priced treatment.

To approach this problem, researchers have developed small affinity domains with the aim to increase tissue penetration while keeping a high specificity to its target. Albumin Binding Domain Derived Affinity Protein (ADAPT) is an example of a small affinity domain of only 7 kDa, which is based on albumin binding domain (ABD) from the streptococcal protein G. Recently, it was shown that the ADAPTs can be further engineered to bind albumin and another relevant target protein of interest simultaneously, which suggests a tolerable half-life in patient serum, alternative administration routes and lower production costs compared to antibody treatments. Furthermore, less side effects are expected due to higher specificity compared to chemotherapy.

This work presents the characterization of novel ADAPT proteins that the target the cancer-related proteins C-C motif ligand 7 (CCL7), vascular endothelial growth factor A (VEGF-A) and carcinoembryonic antigen related cell adhesion molecule 5 (CEACAM5). The new constructs were produced recombinantly in *Escherichia coli (E. coli)* and purified using affinity chromatography. Moreover, the results demonstrate bispecific binding with high affinity towards serum albumin and CCL7 and CEACAM5 respectively, while the ADAPT variants targeting VEGF-A remain to be further developed. Lastly, the importance of different amino acids for structural and binding properties of one CEACAM5 binder are stated. It reveals that the target binding relies on hydrophobic interactions which also can be connected to its poor structural attributes. Accordingly, this project adds new insights about the ADAPTs which can be useful in research towards future clinical applications aimed to improve cancer treatments.



KEY WORDS: cancer therapy, alternative administration routes, ADAPT, ABD, protein technology, albumin, simultaneous bispecificity, CCL7, CEACAM5, VEGF-A

## SAMMANFATTNING

Cancer är fortfarande en av de vanligaste dödsorsakerna världen över och samtidigt finns ett behov av att utöka utbudet av cancerläkemedel som kan användas till terapiresistenta och återkommande cancerformer. Sedan antikroppsbehandling introducerades som cancerbehandling, även kallad immunterapi inom vården, har det varit ett växande fält inom cancerforskning som en mer målinriktad behandling jämfört med cellgifter. Trots stor framgång, kräver immunterapi parenteral administrering dels på grund av begränsad vävnadspenetrering. Om behandlingen administreras intravenöst, behövs specialiserad personal och ingreppet kan även vara obekvämt för patienten. Antikroppsbaserade läkemedel är ofta dyra med anledning av dess kostsamma produktionssteg.

För att lösa det här problemet har små affinitetsdomäner utvecklats med syftet att öka biotillgängligheten och samtidigt behålla en hög specificitet till dess målprotein. Ett exempel på en affinitetsdomän är det albuminbindande domän-deriverade affinitetsproteinet (ADAPT) som är baserad på ett av de albuminbindande domänen i protein G från streptokockbakterier. Nyligen utvecklades bispecifika ADAPT proteiner genom proteinteknik, vilka kan binda albumin och ett annat relevant målprotein samtidigt. Upptäckten tyder på att bispecifika ADAPT kan ha en godtagbar halveringstid i patientserum samt tillåta andra administreringsmetoder och lägre produktionskostnader i jämförelse med antikroppsbehandling. Därtill kan färre biverkningar förväntas på grund av mer målinriktad strategi i jämförelse till cellgifter.

Här presenteras karaktäriseringsdata för nya ADAPT protein mot tre cancerrelaterade målproteiner; C-C motif ligand 7 (CCL7), vascular endothelial growth factor A (VEGF-A) och carcinoembryonic antigen related cell adhesion molecule 5 (CEACAM5). De nya varianterna kan uttryckas rekombinant och produceras i *Escherichia coli (E. coli)* samt renas genom affinitetskromatografi. Projektet genererade bispecifika ADAPT bindare med hög affinitet mot albumin och de respektive målprotein CCL7 och CEACAM5, medan mer utveckling av ADAPT varianterna mot VEGF-A krävs. Slutligen studerades olika aminosyrors betydelse för struktur- och bindingsegenskaper för en CEACAM5 bindare. Detta avslöjade att interaktionen mellan målproteinet och ADAPT mestadels är beroende av hydrofoba interaktioner vilket också kunde kopplas till dess relativa instabilitet. Sammanfattningsvis ger det här projektet nya kunskaper om bispecifika ADAPT protein vilka kan användas i forskning om framtida kliniska applikationer med syfte att förbättra cancerbehandling.



NYCKELORD: cancerbehandling, alternativa administreringsmetoder, ADAPT, ABD, proteinteknik, albumin, simultan bispecificitet, CCL7, CEACAM5, VEGF-A

# TABLE OF CONTENTS

# 1. Introduction

Cancer is the second most common cause of death, accounting for one out of six deaths worldwide during 2018 [1]. Resistance of cancers towards conventional therapies has been indicated as one of the most significant challenges in managing cancer diseases by the National Cancer Institute [2]. The cancer secretome, defined as the macromolecules secreted by cancer cells, has been suggested to be associated with drug resistance and thus provides a set of potential drug targets of novel cancer pharmaceuticals [3]. Membrane proteins are dominating as drug targets of today's pharmaceuticals. They are popular drug targets because of their accessibility on the cell surface and the possibility to modify cell signalling in a targeted manner [4]. Human epidermal growth factor receptor 2 (HER2) is an example of a successful membrane protein drug target in breast cancer [5].

There are various types of cancer therapeutics, including small molecule drugs and biopharmaceuticals, which have their respective characteristics, advantages and limitations. Advantages of small molecule drugs are that they can be taken orally and have a cost-effective production. Disadvantages of small molecule drugs are that they often introduce a systemic effect in the patient due to poor specificity and/or selectivity to target. Small molecules also have a short half-life in patient serum. Antibody therapies exemplify biopharmaceuticals which have a longer circulation time and high selectivity to target. However, antibody therapies require expensive production steps and are often administered intravenously [6,7]. An ideal pharmaceutical would combine the properties of small molecule drugs and antibodies to reduce the side effects and facilitate the distribution and administration of the drug. Small affinity protein domains, or alternative scaffolds, are a result of this idea which enables other administration routes than intravenously, have a cost-effective production and can be engineered to bind specifically to its target. An example of an alternative scaffold is the Albumin Binding Domain Derived Affinity Protein (ADAPT) which are proteins consisting of approximately 46 amino acids. Due to their small sizes, alternative scaffolds have a higher tissue penetration but also a shorter half-life in serum than antibodies [6].

The ADAPT scaffold is based on one of the albumin binding domains (ABD) of the streptococcal protein G. In ADAPTs, an additional binding motif to the ABD motif has been introduced by randomizing 11 amino acids in two out of the three helices of the protein. During the years, the ADAPTs with specificity against numerous targets have been generated through phage display selection libraries, including for example HER2 [8]. ADAPT6 is an ADAPT molecule that targets HER2 and it has shown good performance in clinical studies as an imaging probe and lately also as a possible radionuclide therapy of HER2 positive breast cancer when fused to ABD [9,10]. However, as the short circulation time of the ADAPTs is due to rapid renal clearance, kidney accumulation of the ADAPTs has been observed when not fused to another protein [11]. Consequently, measures to further develop the ADAPT library are needed to ensure safety and efficacy of the ADAPTs without compromising on their small sizes.

Until recently, the ADAPTs did not allow for simultaneous bispecificity due to steric hindering. The HSA binding surface is located in the second and third helix of the protein, and the former ADAPT library was based on randomized amino acids in the first and third helix. However, by moving the randomized amino acid sequence to the first and second helix of the parental ABD035 molecule, ADAPTs can be engineered to be simultaneously specific towards human serum albumin (HSA) and a target protein of interest (*Figure 1*). As albumin is one of the most abundant protein in blood serum and has a half-life of nearly 3 weeks, the simultaneous bispecificity to HSA and a target protein has the potential to increase the half-life of the ADAPTs in patient serum, by taking advantage of the neonatal fragment crystalline receptor recycling (FcRn) mechanism of HSA [11]. In the acidic pH of the endosomes, HSA can bind to FcRn and thus evade degradation in late endosomes. Instead, HSA is recycled and released to the bloodstream again at neutral pH [12]. The half-life of the new generation ADAPTs is hypothesized to be further enhanced by circumventing renal filtration due to increased molecular size in circulation. This discovery motivates the applicability of the new ADAPT proteins as therapeutics that combine the advantages of small molecule drugs and biopharmaceuticals. Thus, further studies of the ADAPTs towards different drug targets could potentially result in new treatment strategies with alternative administration routes [11,12].

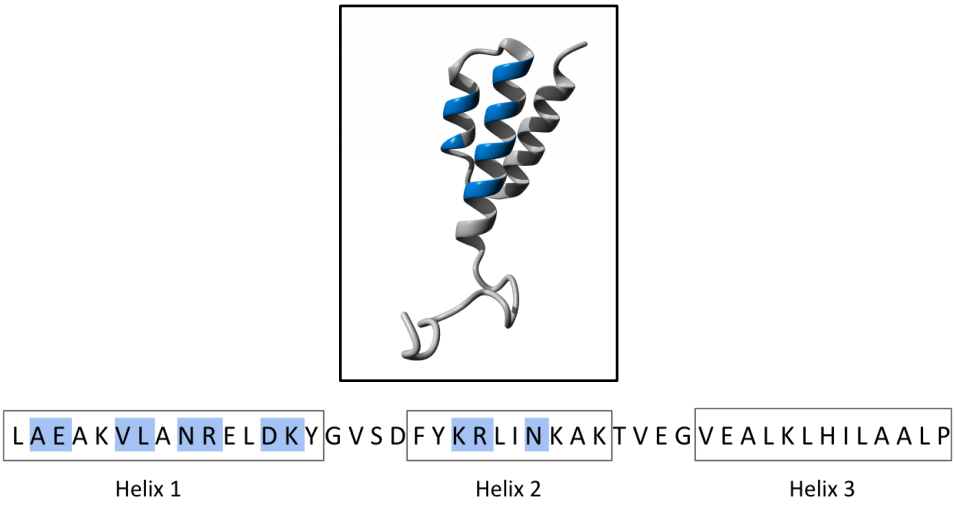



*Figure 1.* ***Randomized positions of the novel ADAPT library.*** *Shows the 11 randomized amino acids in blue of the parental ABD035 molecule used to develop the novel ADAPTs. The image was generated through YASARA (PDB 1GJT).*

## Target proteins

In this work, ADAPTs targeting chemokine C-C motif ligand 7 (CCL7), vascular endothelial growth factor A (VEGF-A) and carcinoembryonic antigen related cell adhesion molecule 5 (CEACAM5) were studied. The target proteins have all been associated with malignancy, severity or invasiveness of cancer diseases [13–15]. CCL7 (UniProt P80098) and VEGF-A (UniProt P15692-4) are secreted by the cells in the tumour microenvironment and by the tumour itself [13,16] while CEACAM5 (UniProt P06731) is a protein overexpressed in the plasma membrane of tumours that also can be cleaved off and released into the bloodstream. Elevated serum CEACAM5 concentration has been observed for cancer patients [17,18]. In figure 2A-C, the structures of the target proteins are illustrated.

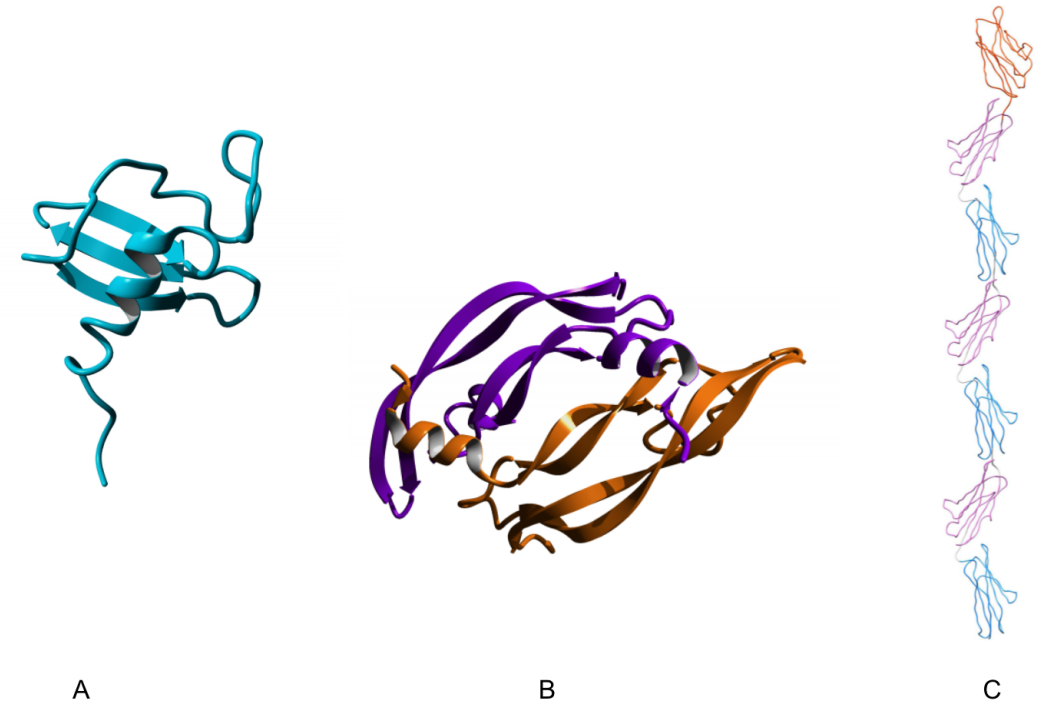



*Figure 2.* ***Structure of target proteins.*** *Shows the structure image of monomeric CCL7 in blue (A), the homodimer VEGF-A, each domain coloured in lilac and orange (B), and CEACAM5 (C). CEACAM5 (C) consists of the variable N domain in red, and the repetitive constant domains A and B, which are coloured in lilac and blue, respectively. The figure was generated using YASARA and structures from NMR and X-ray crystallography for CCL7 (PDB 1NCV) and VEGF-A (PDB 3QTK), respectively, while Chimera and PDB entry 1E07 from homology modelling were used to visualize the amino acid backbone chain of CEACAM5.*

CCL7, also called monocyte chemotactic protein 3 (MCP3), is a 11.2 kDa chemokine (*Figure 2A*) that normally mediates the recruitment of immune cells to the site of infection during immune responses. It has been associated with various disorders in the immune system, like psoriasis. Although its underlying pathways in cancer progression is not fully understood, an increased expression level of CCL7 has been associated with increased tumour metastasis in breast cancer, colorectal cancer and renal cell carcinoma [13,19]. After translation CCL7 is composed of 99 amino acids, which includes a signal peptide of 23 amino acids. The signal peptide is cleaved off upon secretion of CCL7, resulting in a 76 amino acid protein. Although CCL7 is monomeric in solution, it can be dimeric at higher concentrations. CCL7 binds to the C-C motif chemokine receptor (CCR) 1, CCR2, CCR3, CCR5 and CCR10. CCR1 on breast tumour cells, CCR2-positive macrophages, and CCR3 on prostate cells are associated with

tumour angiogenesis, tumour advancement, and tumour invasion, respectively. However, as CCL7 shares its receptors with other chemokines, for example CCL2, the pathological mechanisms of CCL7 are complicated. Furthermore, the evidence regarding the role of CCL7 in cancer is contradicting. Possible tumour promoting effects of CCL7 is suggested to be connected to the overexpression of CCL7 in different cancers. One suggested pathological mechanism of CCL7 is that its expression results in an attractive microenvironment for tumour progression. Nonetheless, CCL7 has in cell studies shown to act as an anti-tumour component by recruiting leukocytes to the tumour and thereby promote an immune response against the tumour [13].

VEGF-A, or VEGF, is a 45 kDa homodimeric glycoprotein (*Figure 2B*) that is overexpressed in many cancers and is associated with malignant tumour growth. The protein is a member of the VEGF platelet-derived growth factor family. There are several isoforms of VEGF, including proteins with a size of 121, 165, 189, 206 amino acids. However, VEGF of 165 amino acids is the most common variant (VEGF165). The 121 isoform is acidic and can thus more frequently diffuse through the cell membrane while the larger isoforms are basic and bind to heparin on the cell surface. VEGF has two corresponding receptors, called VEGF receptor 1 (VEGFR1) and 2 (VEGFR2), where the latter has the role of promoting cell response upon ligand association. VEGF binding to VEGFR1 mainly increases the intracellular signal. In healthy conditions, it facilitates wound healing and plays a significant part in embryonic development by promoting angiogenesis. In cancer diseases, the binding of VEGF to its corresponding VEGF receptor on vascular endothelial cells enables an increased blood vessel supply to the tumour which promotes growth of the tumour. Thus, it is an attractive target for cancer therapies [14]. There are different FDA approved protein-based drugs developed to target VEGF, for example by Bevacizumab, or its receptors VEGFR1 and VEGFR2, for instance Ramucirumab that targets VEGFR2 [20]. Bevacizumab, marketed under Avastin, is a humanized monoclonal antibody approved for metastatic colorectal, lung, kidney and glioblastoma cancers [21], while Ramucirumab is a fully human monoclonal antibody approved in combination therapies for metastatic lung cancer and colorectal cancer [20].

CEACAM5, or CEA, is a protein of 76.8 kDa that is overexpressed in the plasma membrane of colorectal and lung cancer tumours which is proposed to be associated with malignancy and metastasis of cancers [15,17]. There are several proteins in the carcinoembryonic antigen related cell adhesion molecule (CEACAM) family, and CEACAM5 was the first protein to be discovered during the 1960s. The CEACAM proteins have shown diverse functions and relevance in cell adhesion, intracellular signalling, cancer and inflammation, but a common property of the CEACAMs is their large amount of glycosylation. CEACAM5 consists of one N terminal immunoglobulin (Ig)-like variable domain followed by six constant Ig-like domains (*Figure 2C)*. The protein is anchored to the membrane, but it has no intercellular motif [18,22]. The domains are predicted to mainly consist of beta sheets [23], but due to its high glycosylation there have been difficulties to crystallize the full-length protein and study it through X-ray diffraction analysis. Nevertheless, a full-length structure based on homology modelling of each domain separately which was validated using small angle X-ray and neutron scattering analysis revealed a curved orientation of the domains which is the one visualized in Figure 2C [24]. A crystal structure of the variable N fold of CEACAM5 is however available which confirmed the beta sheet structures and showed that the domains are prone to dimerize, proposing a possible CEACAM5-CEACAM5 interaction [25]. Under normal conditions, CEACAM5 is initially expressed during embryonic development and in adults it is present in the colon, stomach, tongue, sweat glands and prostate. The underlying mechanisms of CEACAM5 during tumour development is restricted but it has shown high sensitivity and specificity as a biomarker

and good performance as an overall survival predictor for colon cancer. Suggested mechanisms are that CEACAM5-CEACAM5 interactions between tumours and stromal cells help the tumours to migrate and grow and that its interactions with receptors on immune cells suppress the anti-tumour immune response. Previous clinical trials of CEACAM5 based pharmaceuticals, have investigated the effect of radio conjugated anti-CEACAM5 antibodies [18,22].

## Project outline

The overall aim of this project was to produce and purify 24 novel ADAPT variants against cancer-related secretome and membrane targets that previously have been selected using phage display and then characterize their potential as bispecific binders towards HSA and their targets. The targets which the ADAPTs have been selected against are CCL7, VEGF165 and CEACAM5. The project included small-scale protein production in *Escherichia coli (E. coli),* protein purification using affinity chromatography with an HSA Sepharose matrix and characterization studies. In the characterization study, the structural and oligomeric state of the ADAPTs were determined, respectively, through circular dichroism (CD) spectroscopy and analytical size exclusion chromatography (SEC), in addition to their specificity and affinity towards their target and HSA through surface plasmon resonance (SPR). Furthermore, their simultaneous bispecificity towards their respective target protein and HSA was evaluated. Lastly, the structural and kinetic importance of the randomized amino acids of the most promising ADAPT variant was evaluated through an alanine scan. The alanine scan consisted of replacing the amino acids in the randomized library positions by alanine one by one in the protein sequence. These alanine mutants were produced, purified and characterized according to the stated project workflow (*Figure 3*).

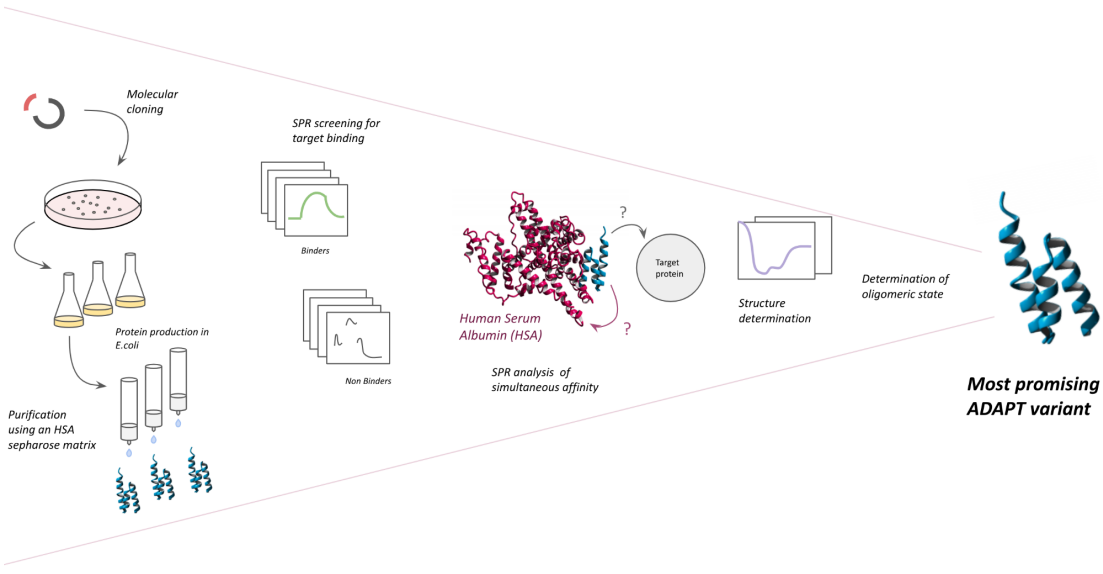



*Figure 3.* ***Project outline.*** *Shows the workflow used in this project. The selected gene inserts were cloned into the production vector, followed by protein production and affinity chromatography purification using an HSA matrix. Then, the variants were screened for target binding followed by analysis of simultaneous binding in SPR. Subsequently, the ADAPT variants were characterized regarding secondary structure, thermal stability and oligomeric state in order to choose the most promising candidate. The amino acids in the randomized positions of the promising candidate were one by one substituted with an alanine and the workflow was used to determine which amino acids that are important for its stability and kinetic properties. The protein structures of HSA and ADAPTs were generated using PDB 1TF0 and YASARA.*

# 2. Materials and Method

## 2.1 Subcloning

All enzymatic reactions were performed according to the manufacturer's instructions, if not stated differently. The PCR purifications, gel extractions and plasmid preparations were performed with QIAGEN's QIAquick PCR purification kit, QIAquick Gel Extraction Kit, and Miniprep protocol, respectively and the resulting DNA concentrations were determined by a NanoDrop 1000 spectrometer (ThermoFisher Scientific).

### 2.1.1 Molecular cloning of ADAPTs targeting CCL7, CEACAM5, and VEGF

The vector which contains AscI and SalI restriction sites, a kanamycin resistance gene and an inducible T7 promoter, was transformed into chemically competent TOP10 *E. coli* cells (made in-house). Genes encoding the 24 ADAPT variants (purchased from Genscript) were PCR amplified to introduce the SalI and AscI restriction sites using Phusion-High fidelity polymerase (ThermoFisher Scientific). The PCR products were analysed through gel electrophoresis using a 1 % agarose gel and subsequently PCR purified. AscI and SalI-HF enzymes, provided by New England Biolabs (NEB), were used to digest the insert and vector DNA. The insert reactions were PCR purified and the vector reactions were gel extracted. A sticky end ligation of the cleaved vector and inserts was performed using T4 DNA ligase (NEB). The ligation reactions were then transformed into TOP10 chemically competent *E. coli* cells using kanamycin plates and PCR screening of selected colonies was performed using DreamTaq DNA polymerase (ThermoFisher Scientific). After colony PCR screening, the sequences of the PCR products were verified using Sanger sequencing (Eurofins).

### 2.1.2 Molecular cloning of alanine mutants

The most interesting variant from the different ADAPT variants, namely $ADAPT_{CEACAM5\_02}$, was analysed further through an alanine scan and by introducing point mutations from $ADAPT_{CEACAM\_01}$ (*Supplementary Table S7*). Primers were designed and ordered from Integrated DNA technologies (IDT) to introduce the desired point mutations in the original $ADAPT_{CEACAM5\_02}$ sequence. A one-step PCR approach was used to introduce the point mutations and the restriction sites Hind-III and EcoRI restriction sites. The gene fragments were cloned into a T7-inducible production vector with carbenicillin resistance and sequenced as described above.

## 2.2 Protein production and purification

Overnight cultures from colonies expressing the correct sequences were plasmid prepped and transformed into BL21 *E. coli* cells (made competent in-house) for production. Subsequently, 5 ml of TSB media with 25 µg/ml kanamycin or 100 µg/ml carbenicillin was inoculated with a single BL21 *E. coli* colony and incubated at 37ºC overnight. 100 ml of fresh TSB media containing 25 µg/ml kanamycin or 100 µg/ml carbenicillin was inoculated with overnight cultures to an $OD_{600}^{START}$ of 0.1. The cultures were cultivated at 150 rpm, 37 ºC until $OD_{600}$ reached 0.75-1, followed by induction through addition of 1 mM Isopropyl β-d-1-thiogalactopyranoside (IPTG). Following that, the cultures were incubated at 25ºC and 150 rpm overnight. The cells were harvested by centrifugation at 4000 rpm and 4ºC for 8 minutes (Rotor

JLA1625, Beckman Coulter). The cell pellet was resuspended in tris-buffered saline Tween-20 buffer (TST; 25mM Tris, 1mM EDTA, 0.2mM NaCl, 0.5 vol% Tween-20) and sonicated using VibraCell (Sonics) with 1 second on/off pulses for 3 minutes and 25 % amplitude.

The cell lysates were centrifuged at 10000xg, 4ºC for 20 minutes (Rotor JA17, Beckman Coulter) followed by 0.45 µm filtration of the supernatant. Subsequently, the lysis samples were purified using affinity chromatography by loading them onto a HSA Sepharose matrix (produced in-house). Elution of the proteins were performed in fractions by adding 0.5M HAc (pH 2.8). The absorbance at 280 nm of the fractions was analysed through spectrophotometry (BioPhotometer Eppendorf), and fractions with absorbance above 0.1 were pooled and freeze-dried overnight using ScanVac CoolSafe freeze dryer (LaboGene). The proteins were then resuspended in phosphate buffer saline (PBS; 1.5 M NaCl, 0.08 M $Na_2HPO_4$, 0.02 M $Na_2HPO_4$, pH 7.5). Due to solvation difficulties, the pH was initially adjusted to 9-10 and when the solution was clear the pH could be adjusted back to approximately 7.5. The purity and yield were estimated with sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) on BioRad's MiniProtean TGX Precast gel, which was stained using GelCode Blue safe protein stain (ThermoFisher Scientific).

## 2.3 Mass spectrometry

The molecular weights of intact protein samples were analysed with electrospray ionisation (ESI) liquid chromatography mass spectrometry (LC/MS). The first eleven samples were prepared as 100 ng/µl dilutions while the majority were analysed as 10µM dilutions in sterilized water. The samples were subsequently analysed in Bruker's Impact II ESI-quadrupole time-of-flight (QTOF) mass analyser connected to a liquid chromatography system with a ProSwift RP-4H Analytical 1x50 mm column (ThermoFisher Scientific).

## 2.4 Circular dichroism

The thermal stability, including refolding capacity and melting temperature, and secondary structure of the pure proteins were analysed through CD spectroscopy with a Chirascan CD spectrometer (Applied Photophysics). The pure protein samples were diluted in PBS to 0.2 mg/ml and analysed in a 0.1 mm path cuvette. The secondary structure was analysed by measuring the circular dichroism of the sample at increasing wavelengths of 195 nm to 260 nm at 20ºC. Then the melting temperature was determined by variable temperature measurements (VTM) of the circular dichroism, or ellipticity, at 221 nm, by increasing the temperature 5º per minute from 20ºC-90ºC, alternatively from 4ºC-90ºC depending on the approximated melting temperature of the protein. The melting temperature was determined by fitting a sigmoid curve to the data points and assessing the temperature at the inflection point of the curves. The refolding capacity of the protein was investigated by analysing the secondary structure after heat treatment at wavelengths 195 nm to 260 nm at 20 ºC.

## 2.5 Size exclusion chromatography

To study the oligomeric state of the most promising ADAPTs, analytical size exclusion chromatography (SEC) on an ÄKTA pure instrument (Cytiva) was performed. All proteins were 0.45 µm filtered and analysed in their present concentrations. 25 µl of samples was injected onto a Superdex 75 increase 50/150 column GL (Cytiva) at a flow rate of 0.45 ml/min. An equilibration volume and elution volume of 6 ml and 4.5 ml was used, respectively, with PBS as a running buffer. Four proteins (Conalbumin, Ovalbumin, Carbonic anhydrase, Ribonuclease A and Aprotinin) with known molecular weights were used as the calibrant.

## 2.6 Surface plasmon resonance

The target binding, cross reactivity and simultaneous affinity were assessed through the Biacore 8K (Cytiva) and Biacore T200 (Cytiva) instruments. In all assays described below, the surfaces were regenerated using 10 mM HCl at a flow of 30 µl/min for 30 s. All protein samples were 0.45 µm filtered and diluted in the running buffer, PBS supplemented with 0.01% Tween-20 (PBST) and kept at 8°C before injection. All analyses were performed in a multicycle manner at 25°C.

The analyses included immobilization of proteins through amine coupling on Sensor S CM-5 surface chips (Cytiva). CCL7 (KTH Secretome group), CEACAM5 (KTH Secretome group) and VEGF165 (Acro Biosystems) were diluted in 10 mM sodium acetate (pH 4.5) to a final concentration of 20 µg/ml, 15 µg/ml and 10 µg/ml, respectively. HSA (Sigma-Aldrich) was diluted in 10 mM sodium acetate (pH 4.5) to final concentrations of 10 µg/ml and 6 µg/ml for immobilization using the 8K instrument and the T200 instrument, respectively. Lastly, 4 µg/ml dilutions in 10 mM sodium acetate (pH 4.5) of the promising anti-VEGF ADAPTs were immobilized on a chip to study the target protein binding capacity, since the VEGF165 surface showed indications of being non-specific. Surface activation was carried out using a 1:1 mixture of 0.4 M ethyl(dimethylaminopropyl) carbodiimide (EDC) and 0.1 M NHS (Cytiva), followed by amine coupling of the ligand and blocking by ethanolamine-HCl pH 8.5 (Cytiva). One flow cell per channel was left as a reference.

### 2.6.1 Screening for target protein and HSA binding and cross reactivity

The pure ADAPT samples were diluted to a concentration of 1000 nM and injected over their respective target proteins and HSA, at 30 µl/min, a contact time of 120 s and dissociation time of 300 s, to screen for target protein and HSA binding. Next, the cross reactivity of the surfaces and the ADAPTs were analysed in the same assay, but by injecting the 1000nM dilutions of the ADAPTs over the target protein surfaces they should not bind to. For the $ADAPT_{VEGF}$ variants, a 1000 nM dilution of VEGF165 was flown over the $ADAPT_{VEGF}$ surfaces to study their direct target binding.

### 2.6.2 HSA capture SPR assay

To screen for simultaneous HSA and target protein binding, 500 nM dilutions of ADAPTs were injected and captured on an HSA surface at 10 µl/min for 120 s, followed by injection of a 500 nM dilution of the target protein at 30 µl/min for 120 s and a dissociation time of 300 s or 600 s. The cross reactivity was further analysed in the HSA capture assay by flowing an irrelevant target protein over the captured ADAPT. To determine the kinetic values, the ADAPTs that showed simultaneous binding were analysed in the same assay again with a 2-fold dilution series of the target protein ranging from 500 nM to 62.5 nM (8K instrument) or 500 nM to 32.25 nM (T200 instrument). The kinetic values were determined by fitting the binding curve assuming a 1:1 interaction between ligand and analyte using Biacore Insight and T200 evaluation software. See Figure 4 for experimental set-up of the HSA capture assay.

### 2.6.3 Dual injection of ADAPTs and CEACAM5

To study the simultaneous HSA and CEACAM5 binding of the ADAPT$_{CEACAM5}$ based mutants, 500 nM dilutions of the mutants were injected at 30 µl/min for 120 s to the HSA surface, followed by injection of 120 s injection of CEACAM5 of varying concentration (125 nM, 250 nM, 500 nM) at 30 µl/min. Dissociation took place at 30 µl/min for 600 s. See Figure 4 for experimental set-up.

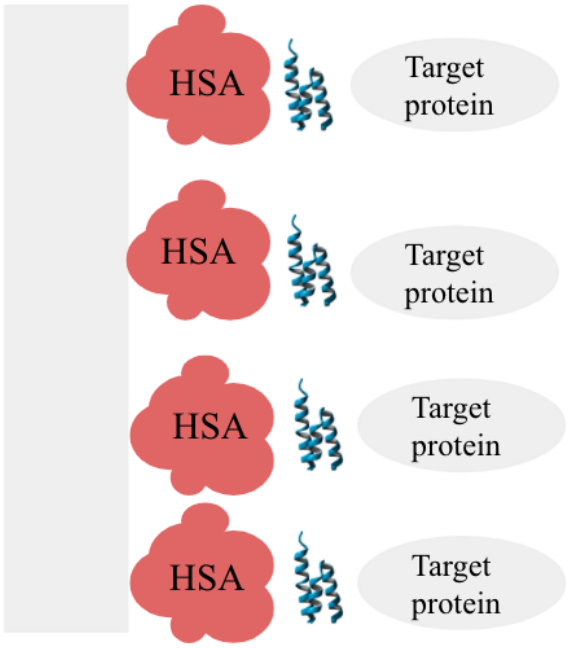


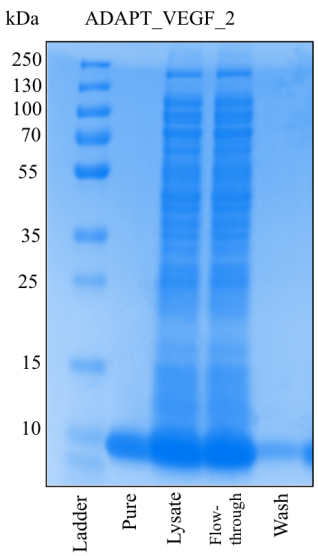

*Figure 4.* ***Set-up of HSA capture SPR assay.*** *Shows the HSA capture SPR assay and the dual injection assay used to study simultaneous binding of the ADAPTs towards HSA and their target proteins. HSA was immobilized on the surface where the ADAPTs were captured. Next, the target protein was flown over the captured ADAPTs. Thus, the ADAPT and the target protein act as ligand and analyte, respectively, in this assay.*

# 3. Results

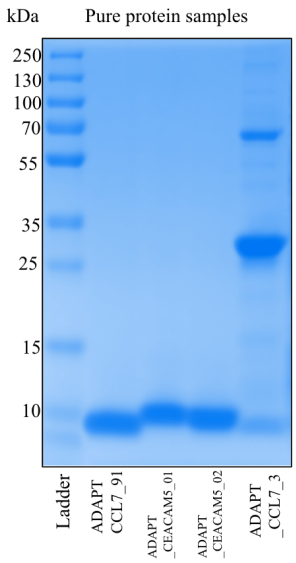
The results of this project have been divided into two parts; (1) Characterization of ADAPTs targeting CCL7, CEACAM5 and VEGF and (2) Alanine scan of ADAPT$_{CECAM5_02}$.

## 3.1 Characterization of ADAPTs targeting CCL7, CEACAM5, and VEGF

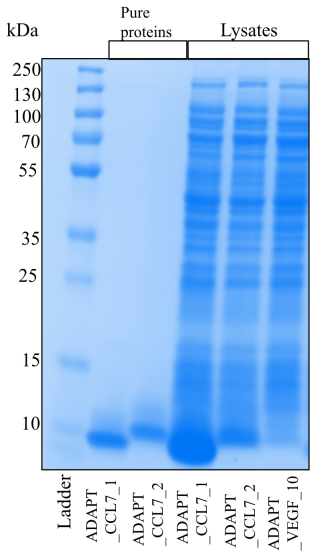
In this part, characterization data of 24 ADAPTs, ten targeting CCL7, six targeting CEACAM5 and eight targeting VEGF, are presented. All ADAPT$_{CCL7}$ and ADAPT$_{CEACAM5}$ were sufficiently produced, ranging from 0.1 to 1.3 milligrams after protein production in 100 ml cultures and purification. The ADAPT$_{VEGF}$ variants had yields of 0.5 to 2 milligrams from a 100 ml production, except for ADAPT$_{VEGF10}$ that only produced 66 µg of protein (*Supplementary Tables S1-S3*). Thus, the purity, yield and approximate molecular weights of all ADAPT variants except for ADAPT$_{VEGF10}$ could be detected and confirmed by SDS PAGE analysis. For three ADAPT$_{CCL7}$ variants, one ADAPT$_{CEACAM5}$ variant and three ADAPT$_{VEGF}$ variants, a lot of protein was detected in the flow-through during purification and for one of the ADAPT$_{CCL7}$ and ADAPT$_{VEGF}$ variants, some protein was detected in the wash sample (*Figure 5A*). The majority of the pure ADAPT samples, with the exception of ADAPT$_{CCL7_03}$, showed a single band at around 7 kDa, which is their expected theoretical molecular weight. ADAPT$_{CCL7_03}$ had three bands visible at around 7 kDa, 30 kDa and 65 kDa (*Figure 5B)*, which could also be identified through MS analysis (*Supplementary Figure S1*). Moreover, the SDS-PAGE gels of the lysate samples illustrate the difference in protein expression of the ADAPTs in this project (*Figure 5C*). As seen in figure 5B and 5C, some proteins travelled slower than expected on the gels, for example ADAPT$_{CEACAM5_01}$ and ADAPT$_{CCL7_02}$, but most of the molecular weights were confirmed to be 0.25 - 1 Da different to their theoretical values according to the LC/MS analysis. However, ADAPT$_{VEGF-7}$ had a molecular weight of 160 Da below its theoretical weight according to the LC/MS. Also, some ADAPTs showed signs of impurities in the MS analysis that were not detected in the SDS-PAGE (*Supplementary Figure S1-S3*).




*A*




*B*




*C*

*Figure 5.* ***SDS-PAGE.*** *Shows SDS-PAGE images loaded with the samples from eluate, lysate, flow-through and wash of ADAPT$_{VEGF-2}$ (A), eluates from ADAPT$_{CCL7-91}$, ADAPT$_{CEACAM5_01}$, ADAPT$_{CEACAM5_02}$ and ADAPT$_{CCL7-3}$ proteins (B), and examples of eluated proteins and lysate samples of ADAPTs targeting CCL7 and VEGF (C). As exemplified in A, the production of some ADAPTs was very efficient which can be seen as a dark band in the lysate sample, although a lot is lost in the flow-through and some in the wash. However, as seen in C, the different ADAPTs showed a varying production efficiency and yield. B illustrates that most proteins showed high purity at the correct molecular weight of around 7 kDa and that ADAPT$_{CCL7-3}$ showed three bands at around 7 kDa, 30 kDa and 60 kDa.*

### 3.1.1 SPR analysis of target binding, cross reactivity, and simultaneous bispecificity

All ADAPT variants were analysed for binding ability towards HSA and the target proteins using SPR where they all showed a strong binding to HSA. VEGF, CEACAM5, CCL7 and HSA were immobilised at a response level of 1970 RU, 815 RU, 790 RU and 6000 RU, respectively for analysis with the 8K instrument. The ADAPTs that showed target binding were also analysed for cross-reactivity to the other target proteins to determine if they are specific towards their respective target. Four ADAPT$_{CCL7}$ variants, three ADAPT$_{CEACAM5}$ variants and none of the ADAPT$_{VEGF}$ were found to be specific towards their respective target protein. Furthermore, the promising variants were investigated for simultaneous bispecificity towards HSA and their target protein which showed that ADAPT$_{CCL7_01}$, ADAPT$_{CCL7_02}$, ADAPT$_{CCL7_03}$ and ADAPT$_{CEACAM5_02}$ had dissociation equilibrium constants ($K_D$) values of $3.88 \cdot 10^{-8}$ M, $6.21 \cdot 10^{-6}$ M, $1.4 \cdot 10^{-7}$ M and $3.64 \cdot 10^{-7}$ M, respectively, to their target proteins when bound to HSA in a capture SPR assay *(Supplementary Table S5 and Figure 6A-D).*

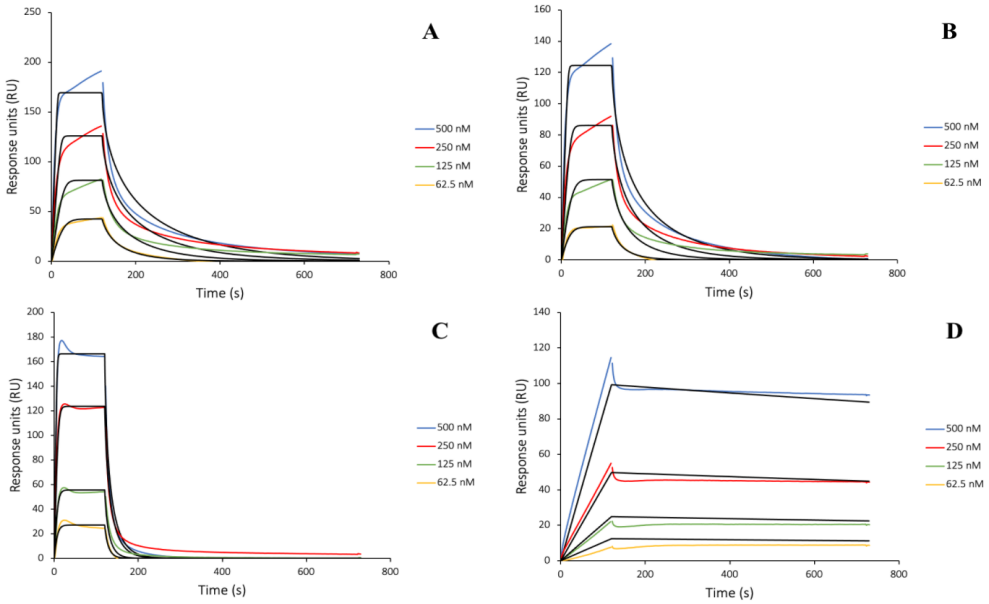



*Figure 6.* ***Simultaneously bispecific binding analysed in HSA capture assay.*** *A-D show blank and reference subtracted SPR sensorgrams and 1:1 binding fits of ADAPT$_{CCL7_01}$ (A), ADAPT$_{CCL7_02}$ (B), ADAPT$_{CCL7_3}$ (C) and ADAPT$_{CEACAM5_02}$ (D). The fitted binding curves, simulating a 1:1 interaction between the ADAPTs and the target proteins, are shown in black while the responses from the 500 nM, 250 nM, 125 nm, and 62.5 nM dilutions of CEACAM5 are shown in blue, red, green and yellow, respectively. The ADAPT$_{CCL7}$ variants showed similar binding curves (A-C), which also is represented by their similar $k_a$ and $k_d$ values. ADAPT$_{CEACAM5_02}$ (D) has a slower on-rate, $k_a$, than the CCL7 binders but a slower off-rate, $k_d$.*

### 3.1.2 Investigation of thermal stability, secondary structure and oligomeric state

The proteins of sufficient amounts were analysed with circular dichroism to study their thermal stability and secondary structure. The most interesting variants from the CD and SPR analysis were subsequently analysed through analytical SEC. Since the anti-VEGF ADAPTs did not perform well in the SPR, they were not included in the SEC analysis. In table 1, the structural data of the most promising ADAPTs are listed, and a complete compilation of all the CD data can be found in Supplementary Table S4 and Figure S4 & S5.

*Table 1.* ***Structural data of most promising ADAPT variants in CD and SPR.*** *Alpha helical structure at 20 °C, melting temperature and the ability to refold after heat denaturation are listed, which were determined by CD. Furthermore, the oligomeric state of the proteins, determined by SEC are stated.*

| ADAPT | Alpha helical structure at 20°C | Melting temp. (°C) | Refolding after heat denaturation | Oligomeric state |
|---|---|---|---|---|
| $ADAPT_{CCL7\_01}$[a] | Yes | 47 | No | Monomeric |
| $ADAPT_{CCL7\_02}$[a] | Yes | 35 | No | Aggregates and monomeric |
| $ADAPT_{CCL7\_03}$[a] | Yes | 32 | Yes | ND |
| $ADAPT_{CEACAM5\_01}$[b] | Yes | 37 | Yes | Monomeric |
| $ADAPT_{CEACAM5\_02}$[b,c] | Yes | 36 | Yes | ND |
| $ADAPT_{CEACAM5\_03}$[b] | Yes | 33 | Yes | ND |
| $ADAPT_{CEACAM5\_04}$[b] | Yes | 72 | Yes | Monomeric |
| $ADAPT_{VEGF\_01}$[a] | Yes | 50 | Yes | ND |
| $ADAPT_{VEGF\_02}$[a] | Yes | 52 | Yes | ND |

[a] VTM program 20-90°C. [b]VTM program 4-90°C. [c]Ran VTM twice. ND - not determined.

The CD analysis showed that 18 out of 22 ADAPTs had alpha helical structure at 20 ºC with melting temperatures ranging from below 30 ºC up to 72 ºC (*Figure 7*). After performing the VTM, 13 ADAPTs could regain their secondary structure while the others were partly or irreversibly denatured (*Figures 8*).

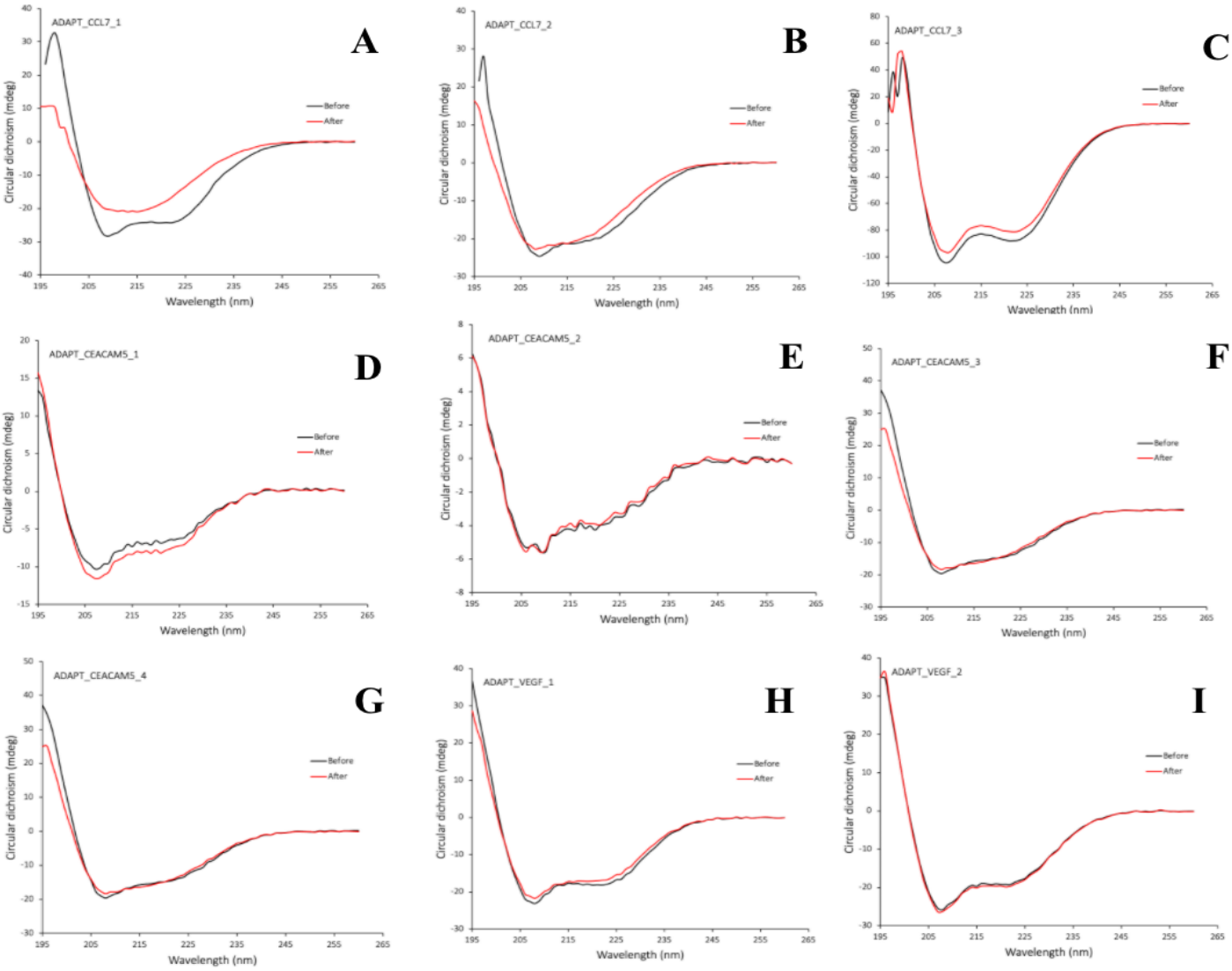



*Figure 7.* ***CD spectrum of ADAPTs.*** *The graphs show the circular dichroism (mdeg) plotted against wavelengths ranging from 195 to 260 (nm) before and after heat treatment of 90 ºC in black and red, respectively. Appearance of alpha helical structure is indicated by dips of the graphs in circular dichroism at approximately 222 and 208 nm. Some ADAPTs showed clear ability to refold which is indicated as the red and black lines overlays (C, D, H & I) while one was irreversibly denatured (A) and one partially denatured (B). For* $ADAPT_{CEACAM5\_03}$ *and* $ADAPT_{CEACAM5\_04}$ *showed slightly unclear CD spectrum (F & G), as they indicated being alpha helical before heat treatment, but their spectrum after heat treatment do not have as clear dips at the desired wavelengths as before.* $ADAPT_{CEACAM5\_02}$ *was determined to be of similar secondary structure before and after heat treatment, but the signal is low and thus does not generate a clear spectrum (E).*

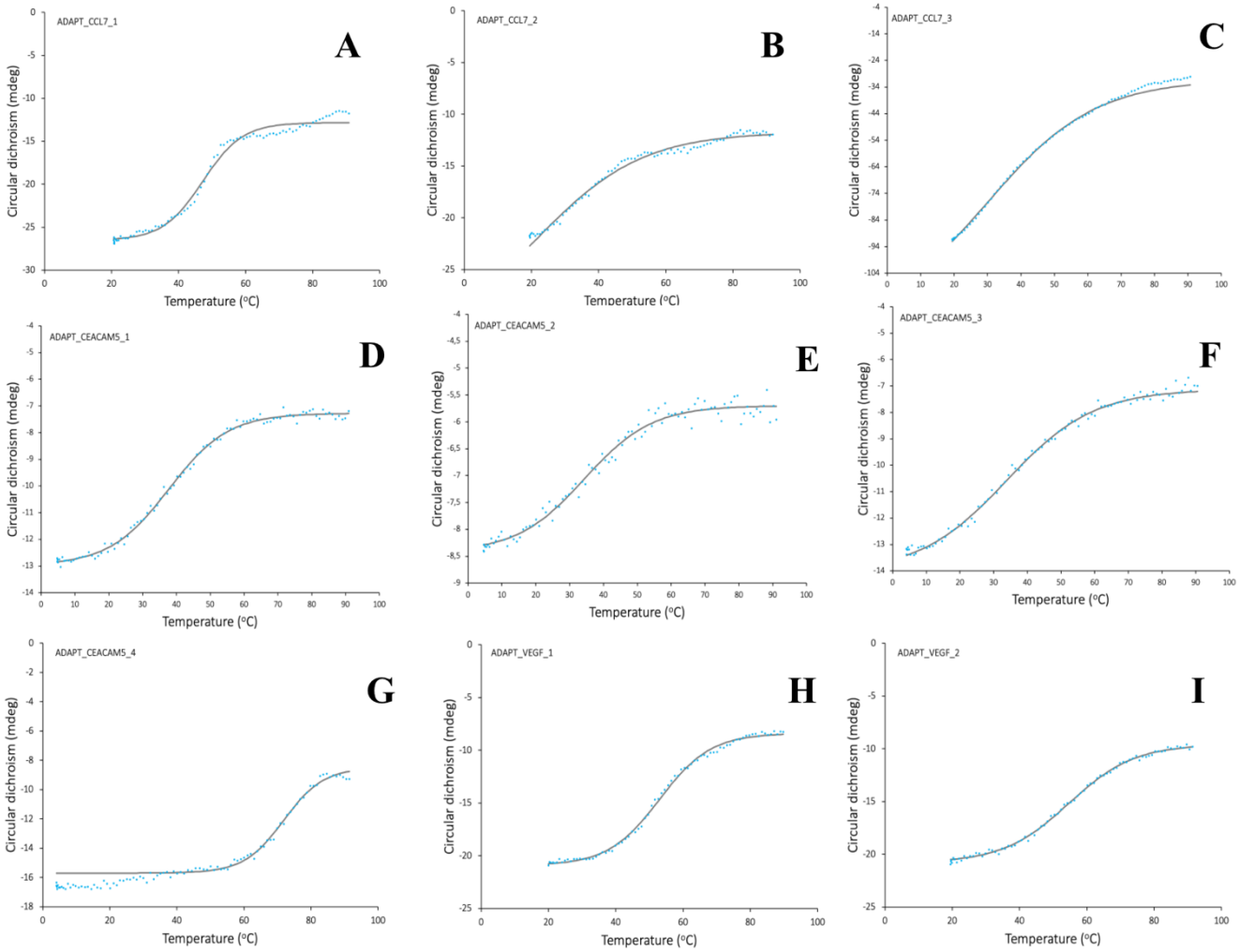



*Figure 8.* ***Melting curves of ADAPTs.*** *The graphs show circular dichroism (mdeg) plotted against temperature (C°) where the measured data points are represented as blue dots and the fitted sigmoid curve as a dark grey line. The melting temperature was determined from these curves, as the temperature at the inflection point of the sigmoidal curve. For D, E, F, G, the measurements were taken from 4 to 90 °C degrees and the others from 20 to 90 °C. Some ADAPTs showed lower melting temperatures of 32-27°C (B-F) and some had higher melting temperatures of 47-72°C (A, G-I).*

The SEC analyses showed that three out of the five analysed ADAPTs were monomeric in solution (*Figure 9A, C, D*). $ADAPT_{CCL7\_02}$ showed to be prone to form aggregates (*Figure 9B*), while the chromatograms of $ADAPT_{CEACAM5\_02}$ and $ADAPT_{CEACAM5\_03}$ had too low absorbance signal to be separated from background noise.

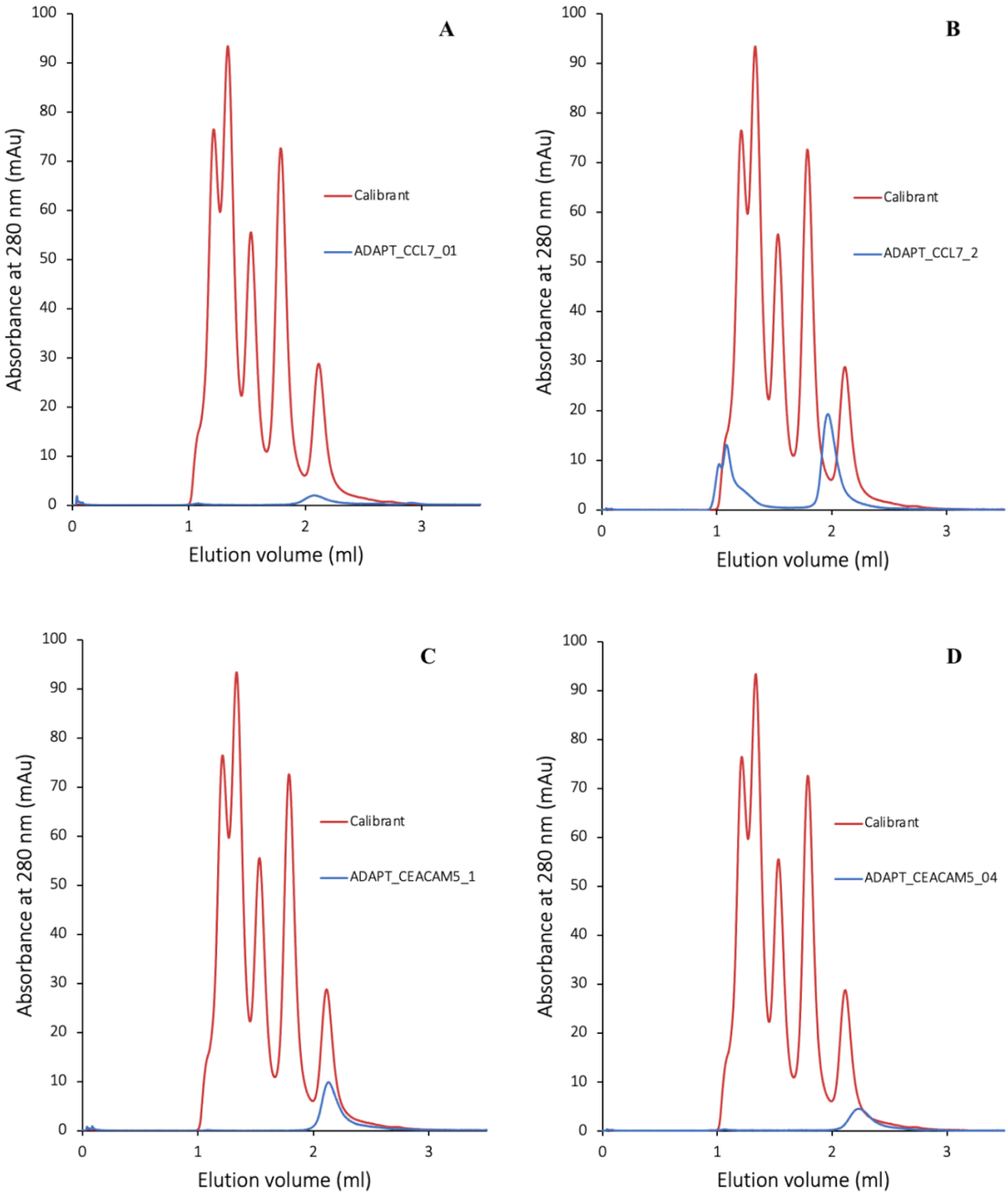



*Figure 9.* ***SEC chromatograms of ADAPTs.*** *The graphs show absorbance at 280 nm plotted against elution volume (ml). The calibrant that was used includes Conalbumin (75 kDa), Ovalbumin (44 kDa), Carbonic anhydrase (29 kDa), Ribonuclease A (13.7 kDa) and Aprotinin (6.5 kDa), where the largest protein is eluted first.* $ADAPT_{CCL7\_01}$, $ADAPT_{CEACAM5\_01}$ *and* $ADAPT_{CEACAM\_04}$ *were thus determined to be monomeric and with correct sizes as their chromatograms showed single peaks eluted at a similar elution volume as Aprotinin (A, C, D) while* $ADAPT_{CCL7\text{-}02}$ *exists as both monomers of 7kDa and aggregates of >75 kDa (B).*

## 3.2 Alanine scan based on ADAPT$_{CEACAM5_02}$

14 ADAPT$_{CEACAM5_02}$ based mutants (M1-M14) were successfully produced and purified with a yield of 0.48 to 2.53 milligrams per 100 ml culture. The original variants, ADAPT$_{CEACAM5_01}$ and ADAPT$_{CEACAM5_02}$ were also produced (*Supplementary Table S6*). A compilation of the characterization in this part of the project is presented below in table 2. In general, four of the mutants (M6-8 and M11) that lost their binding capacity in SPR, were more stable than the original variant. Correspondingly, the mutants that still were simultaneously bispecific (M1-5, M10 and M13), showed worse or the same stability as ADAPT$_{CEACAM5_02}$. The exception was M14, which performed tolerably in both the CD and SPR analysis. All mutants, except for M1 and M4 showed a single band on the SDS-PAGE which corresponded to their theoretical molecular weight of 7 kDa (*Figure 10)*. M1 and M4 showed the same three bands ADAPT$_{CCL7\text{-}03}$ (*Figure 5B)* but these could not be identified as single peaks in the MS analysis. The molecular weights of most mutants were also confirmed in the LC/MS analysis with a 0.25-1 Da difference to their theoretical molecular weights in LC/MS analysis. Some mutants showed signs of impurities in the MS analysis which were not detected in SDS PAGE (*Supplementary Figure S6*). Some of the mutants, M1, M4, M5, M12 and M13 suffered from solubility issues, as they precipitated in pH 7. These proteins were consequently stored in both pH 7 and pH 8.

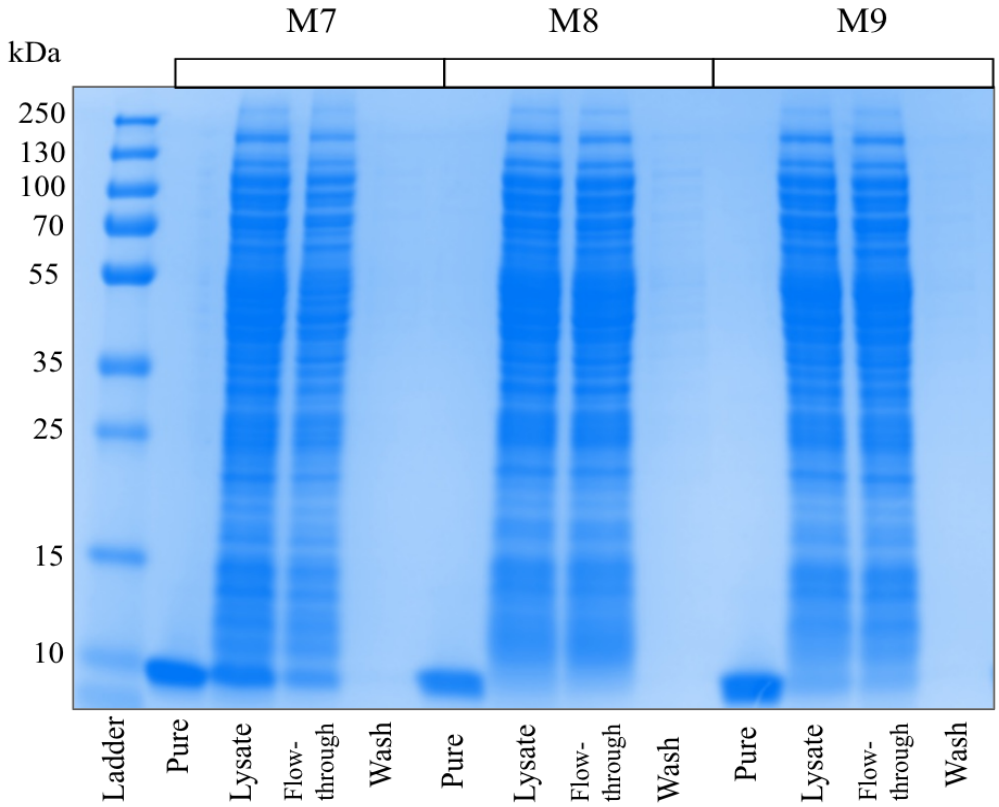



*Figure 10.* ***SDS-PAGE of ADAPT$_{CEACAM5_02}$ mutants M7-M9.*** *The SDS-PAGE illustrates examples of eluated protein, lysate, flow-through and wash samples from three mutants, which are representative of all mutants. Most mutants showed a single band around 7kDa and not a lot of protein in the flow-through and wash. Some lysate samples of the mutants did not show a dark band at 7 kDa, which would indicate a poor production. However, their yields after purification showed an acceptable amounts of proteins.*

*Table 2.* ***Characterization data of ADAPT$_{CEACAM5_02}$ mutants.*** *The table shows the appearance of simultaneous binding, $K_D$ values from SPR analysis, the alpha helical structure at 20°C, melting temperature and the ability to refold after heat denaturation in CD analysis. Some proteins showed alpha helical structure only after VTM, hence the VTM was run twice for those samples.*

| ADAPT | Simultaneous binding to HSA and CEACAM5 | Alpha helical structure at 20°C | Melting temp. (°C) | Refolding after heat denaturation |
|---|---|---|---|---|
| M1 | Yes | No | <30 | N/A |
| M2 | Yes | No (before VTM), Yes (after VTM) | 34 | Yes[b] |
| M3 | Yes | No | <30 | N/A |
| M4 | Yes | No | <30 | N/A |
| M5 | Yes | No | <30 | N/A |
| M6 | No | Yes | 42 | Yes |
| M7 | No | Yes | 42 | Yes |
| M8 | No | Yes | 38 | Yes |
| M9 | No | No | <30 | N/A |
| M10 | Yes | No | 31 | N/A |
| M11 | No | Yes | 43.5 | Yes |
| M12 | Yes | No (before VTM), Yes (after VTM) | 57[a], 39[b] | Yes[b] |
| M13 | Yes | No | <30 | N/A |
| M14 | Yes | Yes | 37 | Yes |
| ADAPT$_{CEACAM_01}$ | No | Yes | 37 | Yes |
| ADAPT$_{CEACAM5_02}$ | Yes | No (before VTM), Yes (after VTM) | 34 | Yes[b] |

[a] *First VTM,* [b] *Second VTM. N/A – not applicable.*

### 3.2.1 Evaluation of the amino acids important for thermal stability, secondary structure and oligomeric state

The mutants showed a varying stability in the CD analysis, six mutants had alpha helical structure at 20 ºC, and five of these could refold after heat treatment, in addition to one of the original variants ADAPT$_{CEACAM5_01}$ (*Figure 11)*. The melting temperatures ranged from 37 to 43.5 ºC (*Table 2 & Supplementary Figure S7)*. The SEC analysis showed that three mutants existed as monomers in solution, even though the SEC chromatogram of mutant 11 showed a peak at an elution volume corresponding to a lower molecular weight than its theoretical mass (*Figure 14*). One of the mutants, mutant 2, gained alpha helical structure after the VTM and thus behaved like the original variant, ADAPT$_{CEACAM5_02}$. For these, the VTM was performed twice, and the proteins showed to be able to refold after the second heat treatment (*Figure 12 & Supplementary Figure S8*). Mutant 12 was stored in both pH 7 and 8 to study the structure stability of the protein since it was difficult to dissolve at neutral pH, and it showed to have an enhanced structural stability in pH 8 compared to pH 7, most likely due to increased solubility (*Figure 13)*. Some of the unstable mutants, M2, M5 and M13 (*Supplementary Figure S9*), precipitated in the cuvette during the CD analysis.

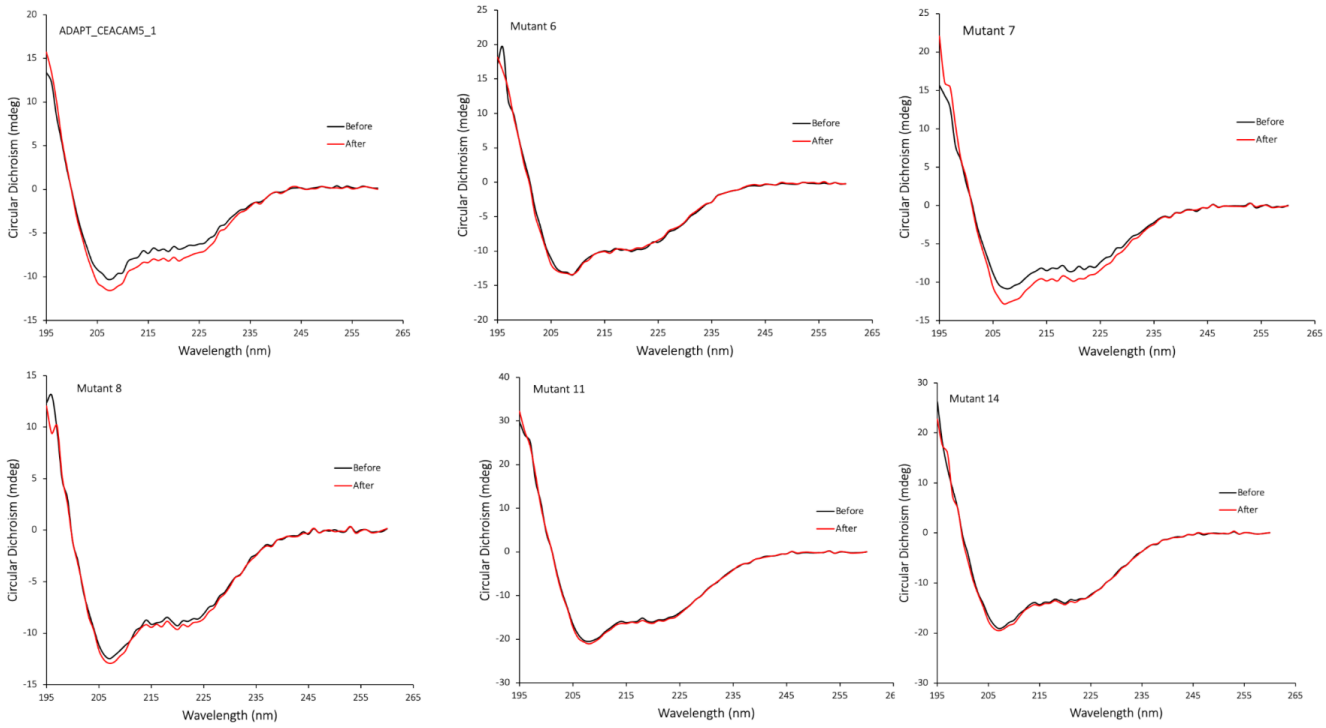



*Figure 11.* ***CD absorbance spectrum of stable mutants and ADAPT$_{CEACAM5_01}$.*** *Graphs show the circular dichroism (mdeg) plotted against wavelengths ranging from 195 to 260 nm before and after heat treatment of 90 ºC in black and red, respectively. Appearance of alpha helical structure is indicated by dips of the graphs in circular dichroism at approximately 222 and 208 nm. These graphs show the mutants that had an alpha helical structure at 20ºC and could successfully refold after heat denaturation.*

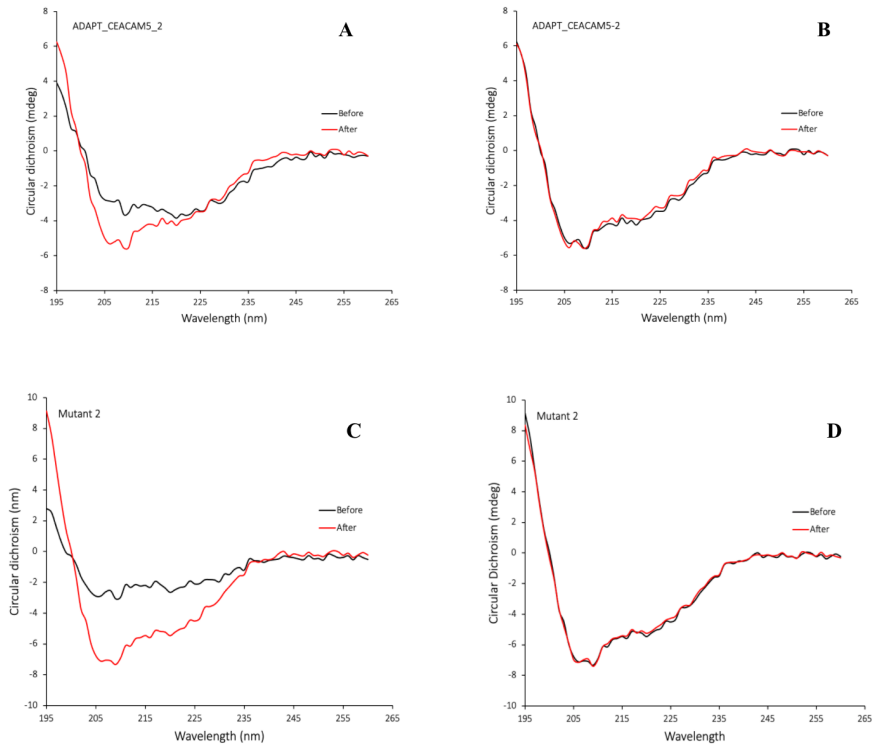



*Figure 12.* ***CD spectrum of ADAPT$_{CEACAM5_02}$ and mutant 2.*** *ADAPT$_{CEACAM5_02}$ and mutant 2 showed a particular behaviour in the CD analysis, as they gained their alpha helical structure after heat denaturation (A, C). After performing an additional heat treatment (4$^{o}$C-90$^{o}$C) heat treatment, they could refold (B, D).*

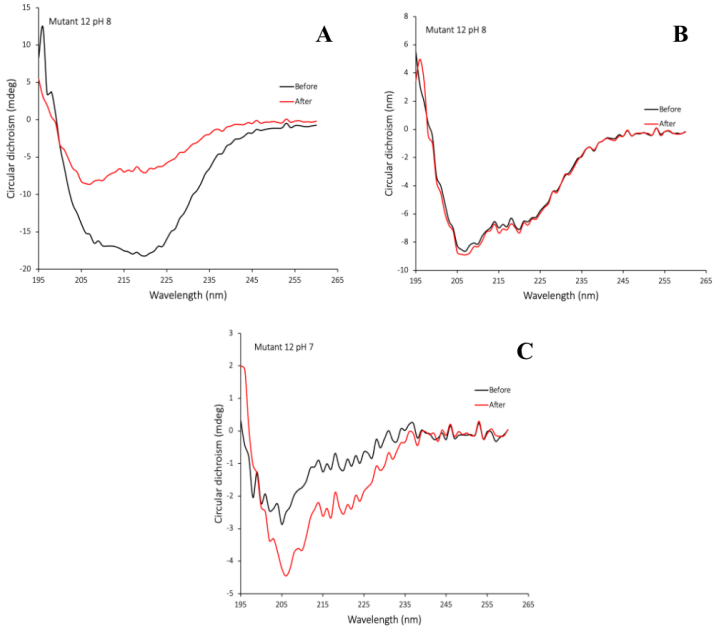



*Figure 13.* ***CD spectrum of mutant 12 at pH 8 and pH 7.*** *To study the structure of proteins that were difficult to dissolve in pH 7, CD analysis was performed on mutant 12 in pH 7 and 8. The mutant 12 in pH 8 got a clearer alpha helical structure after heat treatment in pH 8, although the signal is substantially decreased (A&B). Because of the decrease in signal the data is inconclusive. However, the protein could refold again to the same secondary structure with a similar degree of circular dichroism after an additional heat treatment was performed (B). In pH 7, the mutant 12 did not have alpha helical structure, not even after heat treatment, in addition to very low signal (C).*

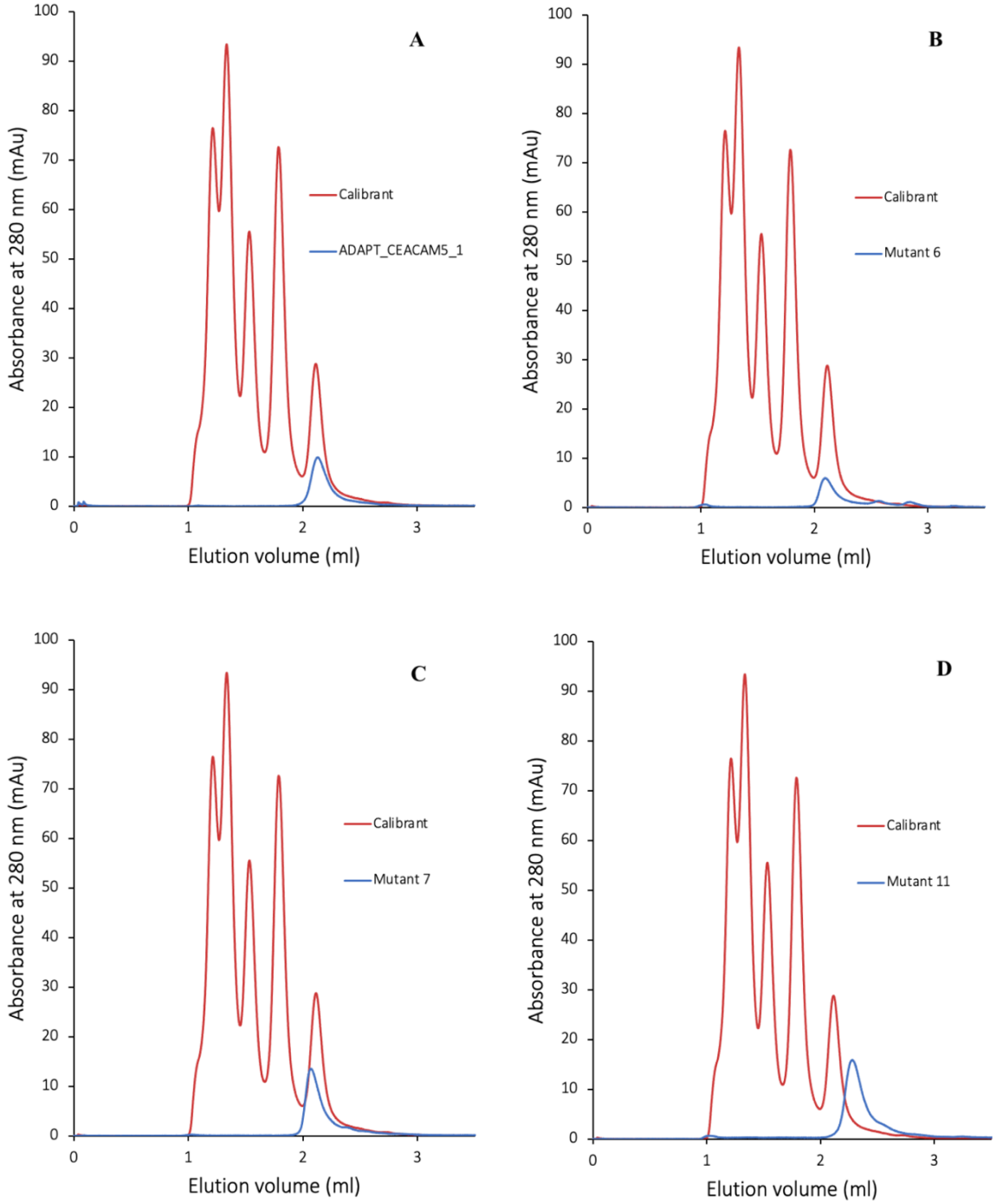



*Figure 14.* ***SEC chromatograms of ADAPT$_{CEACAM5_02}$ based mutants.*** *The graphs show absorbance at 280 nm plotted against elution volume (ml). The calibrant that was used includes Conalbumin (75 kDa), Ovalbumin (44 kDa), Carbonic anhydrase (29 kDa), Ribonuclease A (13.7 kDa) and Aprotinin (6.5 kDa), where the largest protein is eluted first. Mutant 6, mutant 7 and ADAPT$_{CEACAM5_01}$ were determined to be monomeric and with correct sizes as they single peaks eluted at a similar elution volume as Aprotinin (A-C). Mutant 11 showed a single peak at a higher elution volume, which indicates a lower size than its theoretical molecular weight (D).*

### 3.2.2 Evaluation of the amino acids important for simultaneous binding to HSA

HSA was immobilised at response level of 800 RU and 6000 RU for analysis with the T200 instrument and the 8K instrument, respectively. The SPR screening capture assay showed that nine out of 14 mutants were still simultaneously bispecific towards HSA and CEACAM5. The non-binding mutants were still captured to the HSA surface, but their binding to CEACAM5 were impaired (*Figure 15).* In panel J of figure 16, the binding of $ADAPT_{CEACAM5\_02}$ to CEACAM5 was found to be concentration dependent when bound to the HSA surface. The binding curve to HSA, however, distinctly shows a higher on-rate than the binding curve of CEACAM5. For all mutants, the HSA binding capacity was similar to the original variant, which can be seen in the sensorgrams in figure 16. The sensorgrams of the binders (*Figure 16)* and $K_D$ values (*Supplementary Table S8*) were similar to $ADAPT_{CEACAM5\_02}$, with the exception of M5 that had a notably faster off-rate than the others (*Figure 16&17 and Supplementary Table S8*).

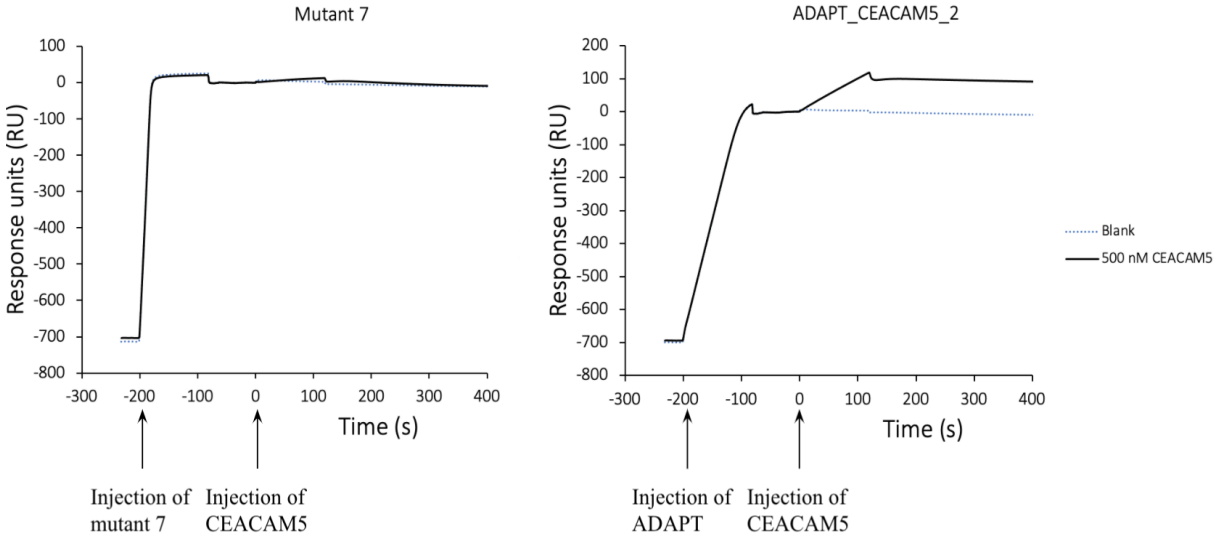



*Figure 15.* ***Screening for simultaneous binding of mutants.*** *The first sensorgram shows mutant 7 captured on the HSA capture, followed by injection of CEACAM5. The sensorgram is representative for all non-binding mutants, compared to the original variant* $ADAPT_{CEACAM5\_02}$ *represented in the sensorgram to the right. The blank cycle is represented as blue dots and the analyte curve is coloured black. As seen in the figure, the blank and analyte cycle completely overlay for mutant 7, while a clear binding curve appears for* $ADAPT_{CEACAM5\_02}$ *when the analyte is injected. However, mutant 7 is still captured by HSA, which is indicated by an increase in response when injected onto the HSA surface.*

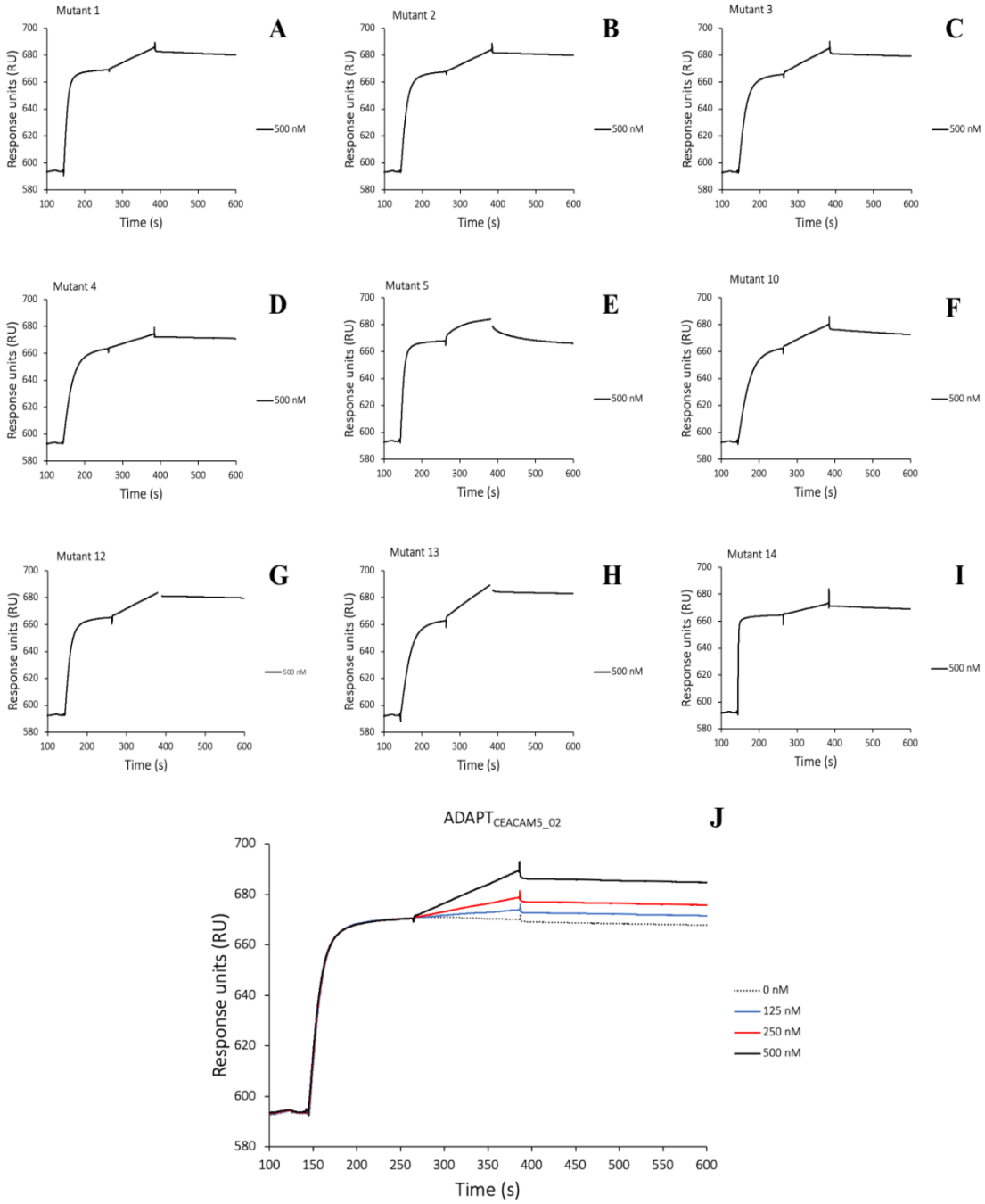



*Figure 16.* ***Binding curves of ADAPT$_{CEACAM5_02}$ based mutants in HSA dual injection assay.*** *The graphs show the ability for simultaneous binding of the ADAPT$_{CEACAM_02}$ based mutants in the dual injection assay, including injection of 500 nM ADAPT at 150 s, followed by injection of 500 nM of the target protein CEACAM5 at 250 s. Most of the mutants, M1-4, M10, M12 and M13 (A-D, F-H) look similar in shape compared to the original variant, seen in J, where CEACAM5 is injected at varying concentrations; 0 nM (dotted black), 125 nM (blue), 250 nM (red) and 500 nM (black). However, the sensorgram of mutant 5 (E) has a distinctly different shape which indicates a faster off-rate than the others. Lastly, the low signal after analyte injection indicated low affinity against CEACAM5 for M4 (D) and M14 (I). The binding to HSA seems to be of similar strength for all mutants.*

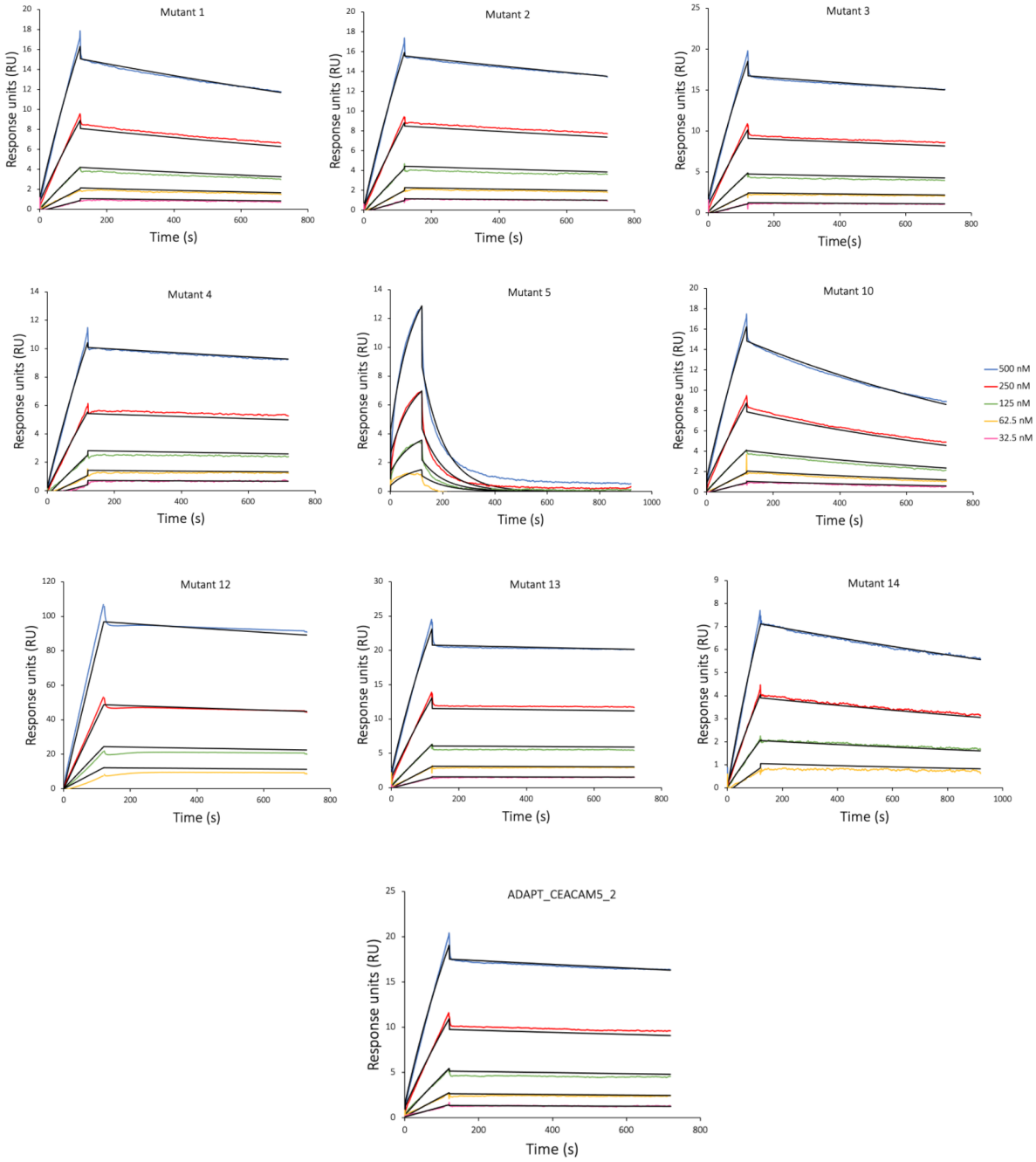



*Figure 17.* ***Binding curves of ADAPT$_{CEACAM5_02}$ based mutants from an HSA capture assay.*** *The graphs show blank and reference subtracted Biacore T200 SPR sensorgrams and fits for ADAPT$_{CEACAM5_02}$ and the simultaneous binding mutants. The fitted binding curves, simulating a 1:1 interaction between the mutants and the target protein, are shown in black while the responses from the 500 nM, 250 nM, 125 nm, 62.5 nM and 31.25 nM dilutions of CEACAM5 are shown in blue, red, green, yellow, and pink, respectively. The graph mutant 12 is from the Biacore 8K SPR instrument and thus its $R_{max}$ cannot be compared to the other graphs.*

# 4. Discussion

Recent studies have shown that small ADAPT proteins of only 7 kDa can be engineered to bind simultaneously to HSA and another target protein of interest [11]. Nevertheless, the bispecific ADAPT proteins remain to be further studied and developed to result in clinical applications. The aim of this master's degree project was to produce 24 novel ADAPT variants targeting cancer-related secreted proteins and membrane proteins that previously had been selected using phage display and then characterize them regarding their potential as simultaneous bispecific binders towards HSA and their targets. Moreover, one of the produced ADAPT molecules, namely $ADAPT_{CEACAM5\_02}$, was considered as the most promising ADAPT variant after the characterization and was thus further analysed in an alanine scan to study the importance of its amino acids for binding and structural properties.

## 4.1 Characterization of ADAPTs targeting CCL7, CEACAM5, and VEGF

Overall, the production and purification scheme of the $ADAPT_{CCL7}$, $ADAPT_{CEACAM5}$ and $ADAPT_{VEGF}$ variants resulted in pure proteins of the correct molecular size (*Figure 5B)*, with some deviations. First, the production efficiency varied across the variants (*Figure 5C)*. Poor production of recombinant protein may be due to many different factors, including for example different plasmid copy numbers of the selected clones [26]. The SDS-PAGE analysis also showed presence of some ADAPT proteins in the wash and flow-through sample (*Figure 5A)*, which can be due to overloading of the column or that a proportion of the produced proteins were misfolded and therefore could not stably bind to the HSA sepharose matrix. For $ADAPT_{CCL7\_03}$, contaminants of 30 kDa and 60 kDa were detected on the SDS-PAGE together with the desired protein of 7 kDa (*Figure 5B)*, which possibly could be a case of *E. coli* chaperonin contamination, for example by the bacterial chaperones GroL/GroS which are of 20-60 kDa and have been identified in different purification schemes [27-29]. A possible solution could be to increase the detergent concentration in the wash buffer or adding denaturation steps before purification to break the chaperon-ADAPT interactions. Moreover, impurities that could not be seen in SDS-PAGE were detected in the LC/MS analysis (*Supplementary Figure S1-S3)*. A possible reason is that small amounts of material that bound non-specifically to the matrix was not sufficiently washed away after sample application in the purification process [30]. One solution to this could be additional washing steps after sample loading. Also, $ADAPT_{CEACAM5\_01}$ and $ADAPT_{CEACAM5\_02}$ include five and three additional positive amino acids in their amino acid sequence, respectively, which could be a probable reason why they travelled slower on the SDS-PAGE gel (*Figure 5B)*. $ADAPT_{VEGF\text{-}7}$ was determined as 160 Da below its theoretical molecular weight in MS analysis which indicated that the protein sequence of the sample is incorrect. This can be due to cleavage of amino acids after translation in the host cell, or, most likely that the incorrect sequence was cloned. Therefore, the interpretation of the characterization regarding this variant is inconclusive.

All ADAPTs showed binding affinity to HSA, both through their ability to be purified by HSA affinity chromatography and by SPR analysis. The parental ABD035 molecule has an affinity to HSA in the femtomolar range [31], and it has been shown that the amino acids in the randomized positions in the ADAPT library does not interfere significantly with this affinity, still being reported to be in the nanomolar range [11]. However, the $K_D$ values against HSA of the new ADAPTs were not determined in this project, which remains to be evaluated in future studies for interesting variants. 18 out of 22 analysed ADAPTs showed presence of alpha helical structure (*Supplementary Table S4 & Figure S4*) which indicates that the ADAPT variants are

prone to be of the desired secondary structure, and their melting temperatures ranged from 30ºC to 67ºC (*Figure 7*). This shows that the changed amino acids sequence of the new ADAPT proteins can result in both decreased and increased thermal stability compared to the parental molecule ABD035, which has a melting temperature of 58 ºC [31]. After heat denaturation, 13 ADAPTs could regain the alpha helical structure (*Figures 8*) which can be beneficial in an eventual large-scale production of ADAPTs.

The two most interesting $ADAPT_{VEGF}$ variants showed promising melting temperatures of 50-52 ºC which can be seen in panel H and I of Figure 7. Although, as mentioned, none of the VEGF variants showed specificity to the target protein VEGF in SPR. As the interaction to VEGF indicated to be unspecific when the VEGF was immobilised on the surface, no conclusion could be made from the primary SPR analysis. Later when the $ADAPT_{VEGF}$ variants were immobilised as ligands, no binding could be seen between the ADAPTs and VEGF. Still, the $ADAPT_{VEGF}$ variants cannot be completely excluded as binders because more parameters remain to be investigated in the surface immobilization of the target protein, in addition to the experimental design where a positive control could be used to verify the results.

Four $ADAPT_{CCL7}$ variants showed to be specific to CCL7 of which three performed well in the HSA capture assay, resulting in $K_D$ values in the micromolar to nanomolar range towards CCL7 when bound to HSA *(Supplementary Table S5 and Figure 6A-C).* Nevertheless, none of these were candidates to be further analysed in this project. Even if $ADAPT_{CCL7\_01}$ showed to be monomeric in SEC analysis (*Figure 9A*) and had a melting temperature of 47 ºC (*Table 1)*, it could not refold after heat denaturation *(Figure 7A).* $ADAPT_{CCL7\_02}$ had a lower melting temperature of 35ºC (*Table 1)*, could refold after heat denaturation *(Figure 7B),* but the SEC analysis showed presence of aggregates (*Figure 9B*). $ADAPT_{CCL7\_03}$ could regain alpha helical structure after heat treatment (*Figure 7C),* the variant had a low melting temperature of 34ºC and had impurities after purification (*Figure 5B*). The contaminants contribute to uncertainties in the data interpretation, as both the $ADAPT_{CCL7\_03}$ and the contaminants can be responsible for the characterization results of this variant. In addition, the fitted binding curves simulating a 1:1 interaction did not overlay well with the experimental data (*Figure 6A-C*), which adds uncertainty to the kinetic values of the anti-CCL7 ADAPTs shown in Supplementary Table S5.

In general, the CD data for the promising $ADAPT_{CEACAM5}$ variants showed similar thermal stability as they could refold after heat treatment (*Figure 7D-G*), despite varying melting temperatures, ranging from 33-72 ºC (*Table 1)*. $ADAPT_{CEACAM5\_04}$ had the most interesting structure characteristics, with a melting temperature of 72ºC (*Table 1)* and determined as monomeric in SEC (*Figure 9D*). Three variants were specific to the target protein CEACAM5, but only $ADAPT_{CECAM5\_02}$ was simultaneously specific to HSA and CEACAM5. However, $ADAPT_{CECAM5\_02}$ had a low melting temperature of 36ºC and was difficult to analyse through SEC because of a low absorbance signal of the resulting peak. $ADAPT_{CECAM5\_01}$ was determined to be monomeric through SEC *(Figure 9C)* and had a marginally higher melting temperature of 37ºC than $ADAPT_{CEACAM5\_02}$ (*Table 1).* The amino acid sequences of $ADAPT_{CECAM5\_02}$ and $ADAPT_{CECAM5\_01}$ only differed by three amino acids while showing quite different characterization results. Therefore, these variants were studied further in an alanine scan to evaluate which amino acids that are important for structure and binding characteristics. Since bispecific ADAPTs are of extra interest in this study, the alanine scan was based on $ADAPT_{CEACAM5\_02}$ even if the variant indicated to suffer from instability issues. Three mutants were also constructed by substituting the amino acids unique for $ADAPT_{CEACAM5\_02}$ to the corresponding position in $ADAPT_{CEACAM5\_01}$.

## 4.2 Alanine scan of ADAPT$_{CEACAM5_02}$

The ADAPT$_{CEACAM5_02}$ based mutants were produced in *E. Coli* and subsequently purified without any major complications *(Figure 10)*, indicating that the studied amino acid substitutions did not have a significant effect on the production properties. Some impurities were observed in SDS PAGE and the LC/MS analysis, explained by the same reasons as described above. A successful HSA affinity chromatography purification with acceptable protein yields implies that the mutants kept their binding to HSA *(Supplementary Table S6)*, which is also confirmed through their response levels in the SPR sensorgrams when captured on HSA (*Figure 15).* Mutants 1, 4, 5, 12 and 13 were especially difficult to dissolve at pH 7. The reasons could be increased isoelectric points of the mutants, insoluble forms of the proteins accumulated in *E. coli* during production [26], or that the solution was saturated due to inaccurate concentration determination by the spectrophotometer. However, the isoelectric points were quite similar across all mutants and the original variants *(Supplementary Table S6*), making the two latter reasons more probable. Analysis of alpha helical structure and thermal stability of mutant 12 were performed at both pH 7 and pH 8 to investigate the pH importance of insoluble mutants at neutral pH, where mutant 12 showed an enhanced structural stability in pH 8 compared to pH 7, most likely due to increased solubility (*Figure 13*).

In figure 18, the proposed important positions for structure and binding of ADAPT$_{CEACAM5_02}$ are illustrated in the ABD structure. When evaluating the structural importance of the different amino acids in ADAPT$_{CEACAM5_02}$, it was found that the amino acids in the randomized position one, three, four, five, nine and ten seem to be of importance for the alpha helical structure of the ADAPT since the mutants without these amino acids, mutant 1, 3, 4, 5, 9 and 10, did not have alpha helical structure at 20ºC (*Table 2)*. These amino acids are uncharged polar, positive and negative in the parental ADAPT$_{CEACAM5_02}$ (*Figure 18A*). Despite their poor alpha helical structure at 20ºC, mutants 1, 3, 4, 5, and 10 were determined as simultaneous binders in SPR analysis at 25ºC. There could be many reasons for this; too low signal of circular dichroism due to inaccurate concentration determination by the spectrophotometer yielding a poor CD spectrum, partly inactive protein in solution, as SEC analysis was difficult to perform it is not known if the proteins exist as monomers or form aggregates, or that the samples were kept at 8ºC until injection in the SPR analysis facilitated protein folding. Also, there is a risk that the ligand-protein interactions are partly unspecific and thus does not require a homogenous protein solution. The amino acids important for binding properties of ADAPT$_{CEACAM5_02}$ were revealed as library positions five, six, seven, eight, nine and 11 which are mostly hydrophobic in addition to one negative and one positive residue *(Figure 18B)*. Two of the amino acids, in randomized position five and nine, are important for both alpha helical structure and simultaneous binding, which respectively are of negative and positive charges. The negative amino acid is located in helix one and the positive amino acid is located in helix two and they appear to point towards each other, which might hold together the structure of the ADAPT through ionic or hydrogen bonds. Library position five is a hydrogen bond acceptor and library position nine is a hydrogen bond donor. Moreover, substitution of the amino acid in randomised position five, resulted in a notably faster off-rate (*Figure 17*). Even if the simultaneous binding towards HSA and CEACAM5 was not impaired, mutant 4 showed a lower response after analyte injection indicating a lower affinity against CEACAM5 when captured on HSA, which could mean that the amino acid in library position 4 is important for binding at some extent in the in the parental ADAPT$_{CEACAM5_02}$. The amino acid was shown to be important for the structure (*Figure 18A*), so an impaired structure might be the reason for its weaker binding to CEACAM5.

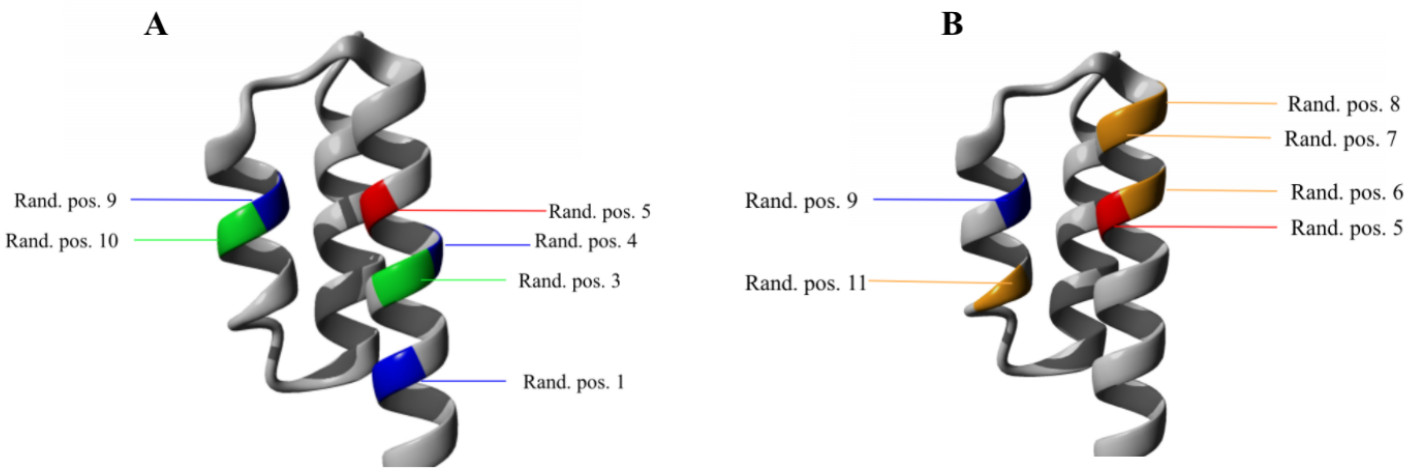



*Figure 18.* ***Important amino acids for alpha helical structure and binding of ADAPT$_{CEACAM5_02}$.*** *The figures show the randomized positions (rand. pos.) in the ADAPT molecule that showed importance for a stable alpha helical structure (A) and for simultaneous binding to CEACAM5 and HSA (B). The positive, negative, uncharged polar and hydrophobic amino acids are coloured blue, red, green and orange, respectively. Worth mentioning is that the positions with hydrophobic amino acids showed in B destabilize the alpha helical structure of ADAPT$_{CEACAM5_02}$, as the mutants without them got improved stability compared to the original variant, yet they are required for the simultaneous binding. Also, by mutating rand. pos. 8 in ADAPT$_{CEACAM5_02}$ to the corresponding amino acid in ADAPT$_{CEACAM5_01}$ (mutant 14), the alpha helical structure was enhanced, which is further proof that the amino acid in ADAPT$_{CEACAM5_02}$ is destabilizing its structure. However, mutant 14 kept its simultaneous binding, indicating that the amino acid in rand. pos. 8 is not the most important amino acid for binding. The images were generated using YASARA and PDB entry 1GJT.*

When mutating library positions six, seven, eight and 11, the alpha helical structure is further enhanced compared to the original variant *(Figure 11)*, which indicates that the original amino acids appear to destabilize the alpha helical structure of the protein. These amino acids are hydrophobic, which indicates that hydrophobicity has a negative effect of the alpha helicity in ADAPTs, especially in the end of the helices. It has been shown that interactions and the type of amino acids in the ends of helices are of great important in the formation and stability of alpha helices in proteins, so-called N-C capping interactions [32]. Proline is known to break alpha helical structure [11], which is a bulky amino acid. Some hydrophobic amino acids are also bulky, which might give rise to breaking of the alpha helical structure. Interestingly, when mutating library position eight in ADAPT$_{CEACAM5_02}$ to the corresponding amino acid of ADAPT$_{CEACAM5_01}$ (mutant 14), the alpha helical structure was also enhanced *(Figure 11)*, which is further proof that ADAPT$_{CEACAM5_02}$ contains a destabilizing amino acid in this position. This library position is also situated in the end of helix one. However, mutant 14 kept its simultaneous binding ability (*Figure 16)*, indicating that library position 8 of ADAPT$_{CEACAM5_02}$ is not the most important amino acid for binding. The particular behaviour of ADAPT$_{CEACAM5_02}$ and mutant 2 in the CD analysis (*Figure 12 & Supplementary Figure S8),* where alpha helical structure is improved after heat denaturation, could be connected to the high surface hydrophobicity content of the proteins. During heat denaturation of the proteins, the hydrophobic interactions are broken. Then, when the denatured proteins are cooled down, it is probably more thermodynamically favourable to fold as alpha helices since they are composed of hydrogen bonds in addition to hydrophobic interactions [33].

The amino acid sequence of $ADAPT_{CEACAM5\_01}$ only differed with three positions compared to $ADAPT_{CEACAM5\_02}$. However, $ADAPT_{CEACAM5\_01}$ showed enhanced stability in both CD and SEC analysis (*Figure 11 & 14*) but lacks the simultaneous binding. However, the findings in this project did not disclose any single mutations that could be responsible for the different characteristics between the two variants. All of the most important amino acids for structure and binding in $ADAPT_{CEACAM5\_02}$ exist in $ADAPT_{CEACAM5\_01}$ except for the amino acid in library position eight. Nonetheless, when library position eight of $ADAPT_{CEACAM5\_02}$ is mutated to the corresponding amino acid in $ADAPT_{CEACAM5\_01}$ the protein does not lose its simultaneous binding, even if a lower $R_{max}$ after capturing the mutant on HSA is observed compared to the other mutants in the dual capture SPR assay (*Figure 16I)*. A proposed reason is that the interplay between the amino acids of $ADAPT_{CEACAM5\_01}$ results in the loss of simultaneous binding, and its different structure properties could be a reason for this. $ADAPT_{CEACAM5\_01}$ also includes more positive charges which might have an impact on the formation of a ternary complex of HSA, ADAPT and CEACAM5. However, an alanine scan of this variant would be necessary to establish which of the amino acids that are most important for its characteristics.

When the oligomeric states of the mutants were analysed, it could be seen that mutation in library position seven (*Figure 14)*, with a hydrophobic amino acid in the original ADAPT, appears to enhance the presence of monomers. However, illustrative SEC data for all mutants and the $ADAPT_{CEACAM5\_02}$ could not be generated, so a comparison is not possible at this stage. For most mutants, too low signals were detected for these to be separated from background absorbance in the SEC chromatograms. The concentrations before and after filtering were analysed to exclude that the protein was filtered away before analysis. One reason could be their relatively low extinction coefficients of 6000 to 7500 $M^{-1}cm^{-1}$ (*Supplementary Table S6*) compared to the extinction coefficient of 30 000 $M^{-1}cm^{-1}$ of the calibrant protein responsible for the highest peak in the chromatogram. Nevertheless, in this case it would be expected to observe the same problem for all mutants. Possible solutions to this could be to optimise the production scheme and produce in larger media volumes to result in more protein, in addition analysis of a more concentrated protein solution.

SPR analysis showed that all mutants could bind to HSA with high affinity (*Figure 15 & 16),* but that some lost their binding to CEACAM5. Their on-rate to HSA appears to be higher than to the target protein. Furthermore, the shapes of the SPR sensorgrams of $ADAPT_{CEACAM5\_02}$ and its mutants (*Figure 6D, 16 & 17*) indicates that the assay suffers from mass transfer limited conditions. The phenomena can be identified by a linear association phase, resulting in a poor curvature of the binding curves and it means that the diffusion from the bulk to the ligand surface is slower than the on-rate, $k_a$ [34]. These non-optimal conditions are suggested to be due to that a proportion of the ADAPTs and/or the target protein is inactive. Aggregation of the ADAPTs cannot be excluded because instability and unpredictable behaviour in CD analysis were observed for these variants (*Figure 12*) and the SEC analysis was difficult to optimise. The spikes of the sensorgrams are most likely due to buffer bulk effects. Also, the $K_D$ value of $ADAPT_{CEACAM5\_02}$ was determined as $3.64 \cdot 10^{-7}$ M from the 8K SPR instrument (Supplementary *Table S5*) and as $1.58 \cdot 10^{-8}$ M from the T200 instrument (*Supplementary Table S8*) which is a huge difference. This can be due to the mass transport limitation and/or poorly fitted curves from the 8K instrument which did not completely overlay the sensorgrams. A reason for a poor fit can be because of the 1:1 interaction assumption which is the base of the fitted curves. It might be that the interaction between analyte and ligand is of another stoichiometric ratio. As CEACAM5 is a relatively large protein of 76.8 kDa compared to the 7 kDa ADAPT, and that CEACAM5 also includes repetitive regions (*Figure 2C*), this is supposed to be a likely scenario. However, despite the sources of uncertainties of the SPR data which

might result in only estimated kinetic values, the results certainly show that the variants can bind to the target protein CEACAM5, and this binding appears to be stable resulting in a slow off-rate $k_d$. Thus, future studies of $ADAPT_{CEACAM5\_02}$ and similar variants are needed to optimise the parameters in the SPR assay to generate true kinetic values. To reduce mass transport limitation, a reduced ligand density or increased flow rate can be applied [34].

### 4.3 Conclusion

In conclusion, this master project has shown that bispecific ADAPTs can be developed towards cancer associated secretome and membrane proteins, but also that further development is required, especially to improve their stability properties. The project generated characterization data for ADAPTs targeting CCL7, CEACAM5 and VEGF and $ADAPT_{CEACAM5\_02}$ was chosen as the most interesting variant to study through an alanine scan. Most likely, the simultaneous binding of $ADAPT_{CEACAM5\_02}$ relies on hydrophobic interactions with a cost of reduced stability. One positive and one negative residue appear to contribute to its alpha helical structure and binding mode by forming hydrogen or ionic bonds between helix one and two and thus stabilising the molecule. The very similar $ADAPT_{CEACAM5\_01}$ protein was shown to be more stable but lacked the simultaneous binding capacity, which could be due to its different structural properties or the additional positive charges of this variant. Although an alanine scan would be necessary to determine the important amino acids for structure and binding in $ADAPT_{CEACAM5\_01}$. From a stability perspective, $ADAPT_{CEACAM5\_04}$ was an interesting variant with a melting temperature of 72 ºC and as it showed to be monomeric in solution. The target protein, CEACAM5 is also of extra interest, as it is an established biomarker for colorectal cancer and its therapeutic accessibility on the cell surface. Other interesting simultaneous binders included $ADAPT_{CCL7\text{-}1}$, $ADAPT_{CCL7\text{-}2}$ and $ADAPT_{CCL7\text{-}3}$, but they suffered from stability or contamination issues, in combination with a target protein of unestablished cancer association. None of the ADAPTs targeting VEGF showed to be specific to the target protein. New phage selections against VEGF could be of interest because of the association of VEGF to cancer malignancies, and that VEGF already is a drug target of approved cancer treatments.

## 5. Future perspectives

This work has resulted in characterization data of new ADAPT proteins targeting proteins that have shown relevance in cancer pathogenesis, namely CCL7, CEACAM5 and VEGF-A, which hopefully will be of interest to researchers that can further develop them towards new cancer pharmaceuticals. The ultimate purpose of developing ADAPTs for cancer-related targets is to provide health care with alternative cancer treatments that allows for new administration routes and less side effects which means less suffering for patients. The $ADAPT_{CEACAM5}$ variants should be of most interest for further development towards modern cancer pharmaceuticals that facilitate distribution, increase half-life in patients and lower production cost compared to immunotherapies based on antibodies. Using the data of this project, it is possible to develop improved $ADAPT_{CEACAM5}$ binders by designing a maturation library where the number of randomized positions is reduced and with less variation followed by new selections against CEACAM5. Furthermore, the characterization information can be used to develop a new bispecifc ADAPT library to be used in new phage selections, to increase the change of selecting binders with increased structural stability and also to include new target proteins.

# 6. Acknowledgements

I would first like to point my acknowledgments to Sophia Hober that made this project possible. Thank you for letting me be a part of such an inspiring research field and for welcoming me in your research group. Next, I want to thank my supervisor Marit Möller who also made the previous phage selections of these ADAPT variants. I really appreciate your excellent supervising in both academic writing and laboratory work and your uplifting feed-back. I also want to thank Sarah Lindbo for your helpful comments on this thesis. Acknowledgements are also given to the KTH Secretome group for providing me with CEACAM5 and CCL7 proteins. Lastly, I would like to expand my acknowledgments to all members of the Hober lab and everyone at Floor 3 for interesting discussions which helped in project implementation and brought new ideas to the project.

# 7. References


1. World Health Organization (WHO). Cancer [Internet]. [cited 2020 Sep 19]. Available from: https://www.who.int/health-topics/cancer#tab=tab_1

2. Cancer Drug Resistance - National Cancer Institute [Internet]. [cited 2020 Dec 16]. Available from: https://www.cancer.gov/research/annual-plan/scientific-topics/drug-resistance

3. Patel S, Ngounou Wetie AG, Darie CC, Clarkson BD. Cancer secretomes and their place in supplementing other hallmarks of cancer. Adv Exp Med Biol. 2014;806:409–42.

4. Yin H, Flynn AD. Drugging membrane protein interactions. Annu Rev Biomed Eng. 2016;18:51–76.

5. Wang J, Xu B. Targeted therapeutic options and future perspectives for HER2-positive breast cancer. Signal Transduct Target Ther. 2019;4:34.

6. Nilvebrant J, Astrand M, Löfblom J, Hober S. Development and characterization of small bispecific albumin-binding domains with high affinity for ErbB3. Cell Mol Life Sci. 2013;70:3973–85.

7. Reynolds T, de Zafra C, Kim A, Gelzleichter TR. Overview of Biopharmaceuticals and Comparison with Small-molecule Drug Development. Nonclinical development of novel biologics, biosimilars, vaccines and specialty biologics. Elsevier; 2013. p. 3–33.

8. Nilvebrant J, Åstrand M, Georgieva-Kotseva M, Björnmalm M, Löfblom J, Hober S. Engineering of bispecific affinity proteins with high affinity for ERBB2 and adaptable binding to albumin. PLoS ONE. 2014;9:e103094.

9. Garousi J, von Witting E, Borin J, Vorobyeva A, Altai M, Vorontsova O, et al. Radionuclide therapy using ABD-fused ADAPT scaffold protein: Proof of Principle. Biomaterials. 2021;266:120381.

10. Bragina O, von Witting E, Garousi J, Zelchan R, Sandström M, Orlova A, et al. Phase I Study of 99mTc-ADAPT6, a Scaffold Protein-Based Probe for Visualization of HER2 Expression in Breast Cancer. J Nucl Med. 62:493–9.

11. von Witting E, Lindbo S, Lundqvist M, Möller M, Wisniewski A, Kanje S, et al. Small Bispecific Affinity Proteins for Simultaneous Target Binding and Albumin-Associated Half-Life Extension. Mol Pharm. 2020;

12. Sand KMK, Bern M, Nilsen J, Noordzij HT, Sandlie I, Andersen JT. Unraveling the Interaction between FcRn and Albumin: Opportunities for Design of Albumin-Based Therapeutics. Front Immunol. 2014;5:682.

13. Liu Y, Cai Y, Liu L, Wu Y, Xiong X. Crucial biological functions of CCL7 in cancer. PeerJ. 2018;6:e4928.

14. Carmeliet P. VEGF as a key mediator of angiogenesis in cancer. Oncology. 2005;69 Suppl 3:4–10.

15. Blumenthal RD, Hansen HJ, Goldenberg DM. Inhibition of adhesion, invasion, and metastasis by antibodies targeting CEACAM6 (NCA-90) and CEACAM5 (Carcinoembryonic Antigen). Cancer Res. 2005;65:8809–17.

16. Goel HL, Mercurio AM. VEGF targets the tumour cell. Nat Rev Cancer. 2013;13:871–82.

17. Expression of CEACAM5 in cancer - Summary - The Human Protein Atlas [Internet]. [cited 2021 Feb 4]. Available from: https://www.proteinatlas.org/ENSG00000105388-

CEACAM5/pathology

18. Conaghan P, Ashraf S, Tytherleigh M, Wilding J, Tchilian E, Bicknell D, et al. Targeted killing of colorectal cancer cell lines by a humanised IgG1 monoclonal antibody that binds to membrane-bound carcinoembryonic antigen. Br J Cancer. 2008;98:1217–25.

19. CCL7 protein expression summary - The Human Protein Atlas [Internet]. [cited 2021 Mar 9]. Available from: https://www.proteinatlas.org/ENSG00000108688-CCL7

20. Lacal PM, Graziani G. Therapeutic implication of vascular endothelial growth factor receptor-1 (VEGFR-1) targeting in cancer cells and tumour microenvironment by competitive and non-competitive inhibitors. Pharmacol Res. 2018;136:97–107.

21. Avastin (bevacizumab) Information | FDA [Internet]. [cited 2021 May 24]. Available from: https://www.fda.gov/drugs/postmarket-drug-safety-information-patients-and-providers/avastin-bevacizumab-information

22. Beauchemin N, Arabzadeh A. Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis. Cancer Metastasis Rev. 2013;32:643–71.

23. Bates PA, Luo J, Sternberg MJ. A predicted three-dimensional structure for the carcinoembryonic antigen (CEA). FEBS Lett. 1992;301:207–14.

24. Boehm MK, Perkins SJ. Structural models for carcinoembryonic antigen and its complex with the single-chain Fv antibody molecule MFE23. FEBS Lett. 2000;475:11–6.

25. Korotkova N, Yang Y, Le Trong I, Cota E, Demeler B, Marchant J, et al. Binding of Dr adhesins of Escherichia coli to carcinoembryonic antigen triggers receptor dissociation. Mol Microbiol. 2008;67:420–34.

26. Fakruddin M, Mohammad Mazumdar R, Bin Mannan KS, Chowdhury A, Hossain MN. Critical Factors Affecting the Success of Cloning, Expression, and Mass Production of Enzymes by Recombinant E. coli. ISRN Biotechnol. 2013;2013:590587.

27. Hartl FU, Bracher A, Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. Nature. 2011;475:324–32.

28. Rohman M, Harrison-Lavoie KJ. Separation of copurifying GroEL from glutathione-S-transferase fusion proteins. Protein Expr Purif. 2000;20:45–7.

29. Yantsevich AV, Dzichenka YaV, Ivanchik AV, Shapiro MA, Trawkina M, Shkel TV, et al. Proteomic analysis of contaminants in recombinant membrane hemeproteins expressed in E. coli and isolated by metal affinity chromatography. Appl Biochem Microbiol. 2017;53:173–86.

30. Danielsson Å. Affinity Chromatography. Biopharmaceutical Processing. Elsevier; 2018. p. 367–78.

31. Jonsson A, Dogan J, Herne N, Abrahmsén L, Nygren P-A. Engineering of a femtomolar affinity binding protein to human serum albumin. Protein Eng Des Sel. 2008;21:515–27.

32. Doig AJ, Baldwin RL. N- and C-capping preferences for all 20 amino acids in alpha-helical peptides. Protein Sci. 1995;4:1325–36.

33. Makhatadze GI. Thermodynamics Of α-Helix Formation. Peptide Solvation and H-Bonds. Elsevier; 2005. p. 199–226.

34. Marquart JA. Chapter 4. sprpages – getting a feeling for the curves. In: Schasfoort RBM, editor. Handbook of surface plasmon resonance. Cambridge: Royal Society of Chemistry; 2017. p. 106–48.

# Supplementary Tables and Figures

## Production & purification data of ADAPTs

*Table S1.* ***Anti-CCL7 ADAPT variants.***

| Construct name | Yield (mg) | Theoretical Mw (Da) | Isoelectric point | Extinction coefficient | MS Determined Mw (Da) |
|---|---|---|---|---|---|
| ADAPT_CCL7_01 | 0.650 | 7008.84 | 5.68 | 15470 | 7008.4605 |
| ADAPT_CCL7_02 | 0.594 | 6974.78 | 5.45 | 13980 | 6974.4535 |
| ADAPT_CCL7_03 | 0.949 | 7174.96 | 5.90 | 18450 | 7108.6156, 30706.284, 61348.3734 |
| ADAPT_CCL7_05 | 0.792 | 7033.94 | 5.90 | 13980 | 7033.5438 |
| ADAPT_CCL7_09 | 0.132 | 6983.88 | 5.90 | 11460 | 6983.5232 |
| ADAPT_CCL7_14 | 0.207 | 6978.83 | 5.68 | 13980 | 6978.433 |
| ADAPT_CCL7_18 | 0.625 | 6944.80 | 5.69 | 13980 | 6944.473 |
| ADAPT_CCL7_20 | 0.605 | 7080.98 | 5.47 | 13980 | 7080.4409 |
| ADAPT_CCL7_21 | 0.807 | 7031.85 | 5.46 | 9970 | 7031.4091 |
| ADAPT_CCL7_91 | 0.403 | 6983.88 | 5.90 | 11460 | 7004.5438 |

*Table S2.* ***Anti-CEACAM5 ADAPT variants.***

| Construct name | Yield (mg) | Theoretical Mw (Da) | Isoelectric point | Extinction coefficient | MS Determined Mw (Da) |
|---|---|---|---|---|---|
| ADAPT_CEACAM5_01 | 1.394 | 7137.03 | 6.57 | 5960 | 7136.5994 |
| ADAPT_CEACAM5_02 | 1.307 | 7136.0 | 6.45 | 7450 | 7135.6461 |
| ADAPT_CEACAM5_03 | 0.926 | 7066.97 | 5.68 | 9970 | 7066.5378 |
| ADAPT_CEACAM5_04 | 1.192 | 7016.94 | 6.45 | 9970 | 7016.6043 |
| ADAPT_CEACAM5_08 | 1.35 | 7119.94 | 6.35 | 9970 | 7119.5859 |
| ADAPT_CEACAM5_22 | 0.600 | 7172.10 | 6.38 | 15470 | Impure |

*Table S3.* ***Anti- VEGF ADAPT variants.***

| Construct name | Yield (mg) | Theoretical Mw (Da) | Isoelectric point | Extinction coefficient | MS Determined Mw (Da) |
|---|---|---|---|---|---|
| ADAPT_VEGF_01 | 0.480 | 6843.69 | 5.68 | 2980 | 6843 |
| ADAPT_VEGF_02 | 1.99 | 7070.94 | 6.02 | 15470 | 7070.5864 |
| ADAPT_VEGF_04 | 0.62 | 6755.57 | 5.90 | 2980 | 6755 |
| ADAPT_VEGF_06 | 2.43 | 7070.91 | 5.75 | 4470 | 7070.4889 |
| ADAPT_VEGF_07 | 0.1231 | 7159.07 | 5.81 | 19480 | 6999.5715 |
| ADAPT_VEGF_10 | 0.00662 | 7159.07 | 5.81 | 19480 | ND |
| ADAPT_VEGF_23 | 0.828 | 7105.94 | 5.68 | 19480 | 7105 |
| ADAPT_VEGF_42 | 0.553 | 6984.95 | 5.90 | 13980 | Impure |

## MS chromatogram of ADAPTs

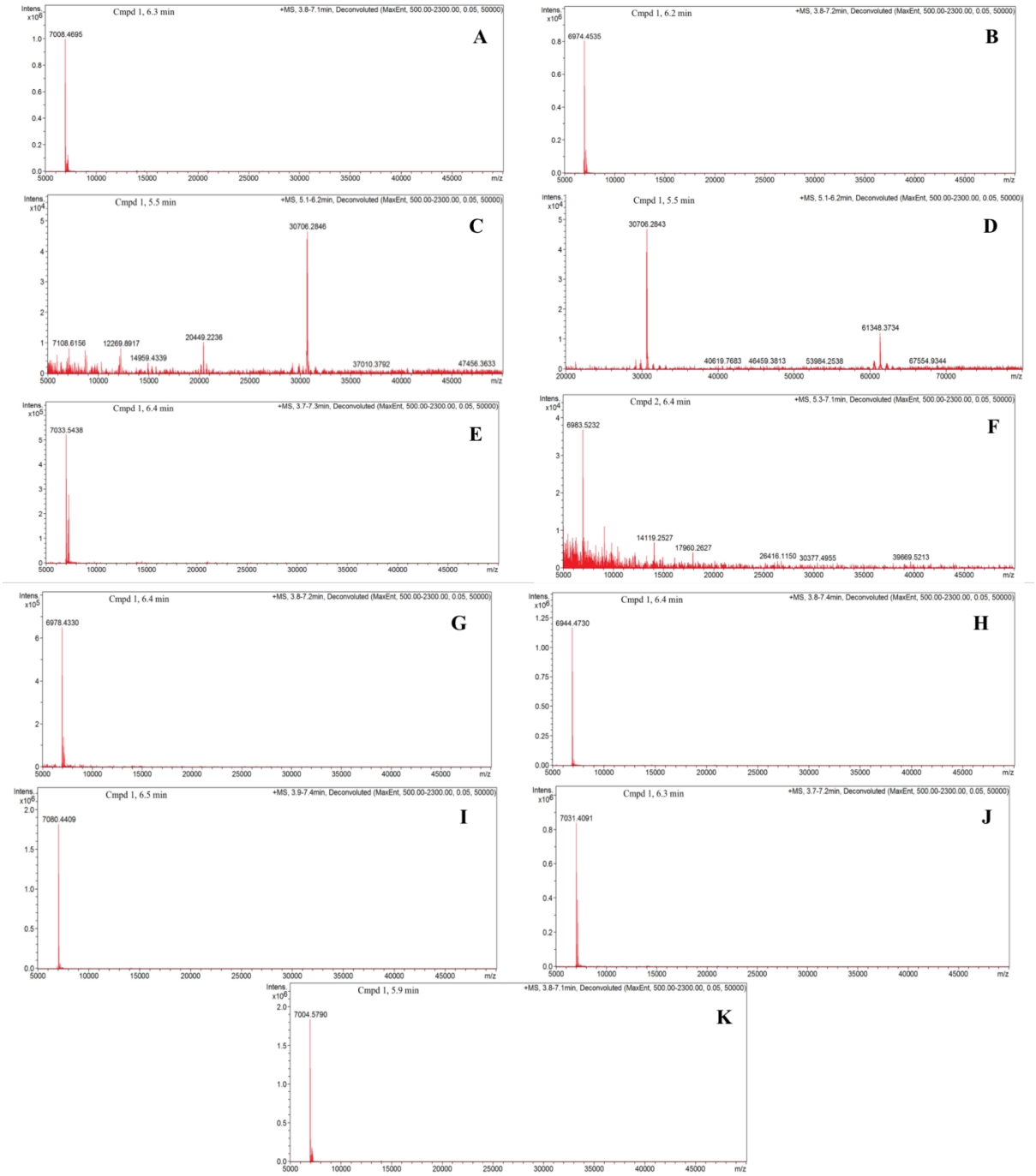



*Figure S1.* ***MS results of ADAPT$_{CCL7}$ variants.*** *Shows the LC/MS chromatogram of ADAPT$_{CCL7-1}$ (A), ADAPT$_{CCL7-2}$ (B), ADAPT$_{CCL7-3}$ (C & D), ADAPT$_{CCL7-5}$ (E), ADAPT$_{CCL7-9}$ (F), ADAPT$_{CCL7-14}$ (G), ADAPT$_{CCL7-18}$ (H), ADAPT$_{CCL7-20}$ (I), ADAPT$_{CCL7-21}$ (J), ADAPT$_{CCL7-91}$ (K).*

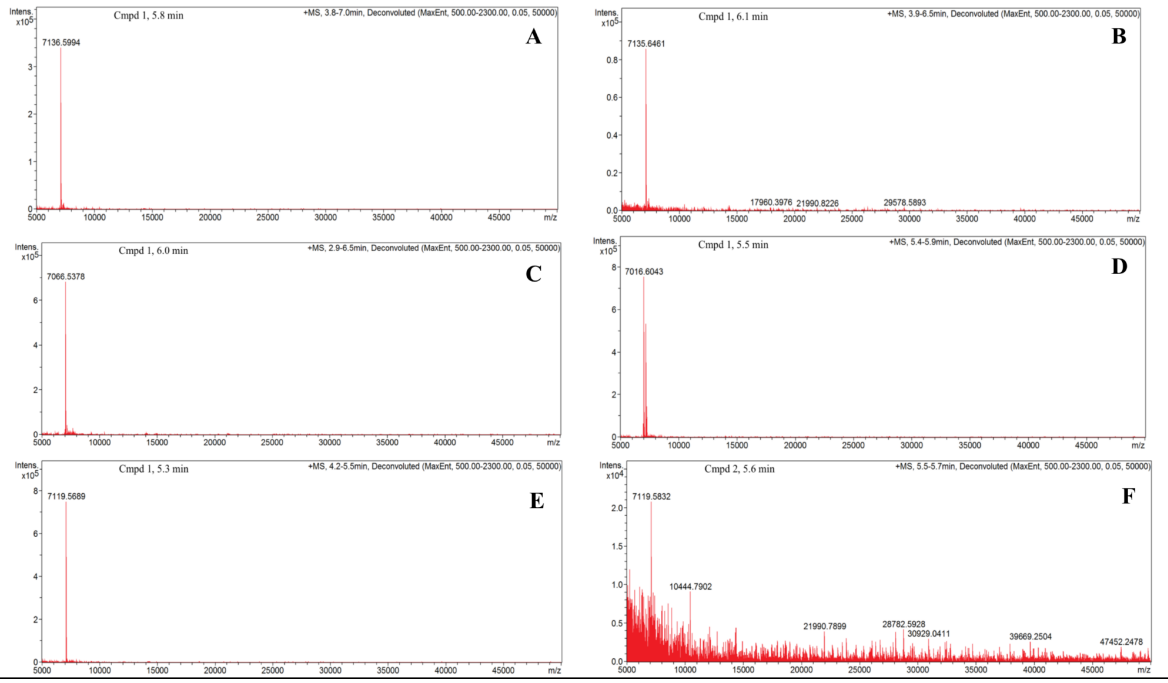



*Figure S2.* ***MS results of ADAPT$_{CEACAM5}$ variants.*** *Shows the LC/MS chromatograms of ADAPT$_{CEACAM5_01}$ (A), ADAPT$_{CEACAM5_02}$ (B), ADAPT$_{CEACAM5_03}$ (C), ADAPT$_{CEACAM5_04}$ (D), ADAPT$_{CEACAM5_08}$ (E), ADAPT$_{CEACAM5_22}$ (F).*

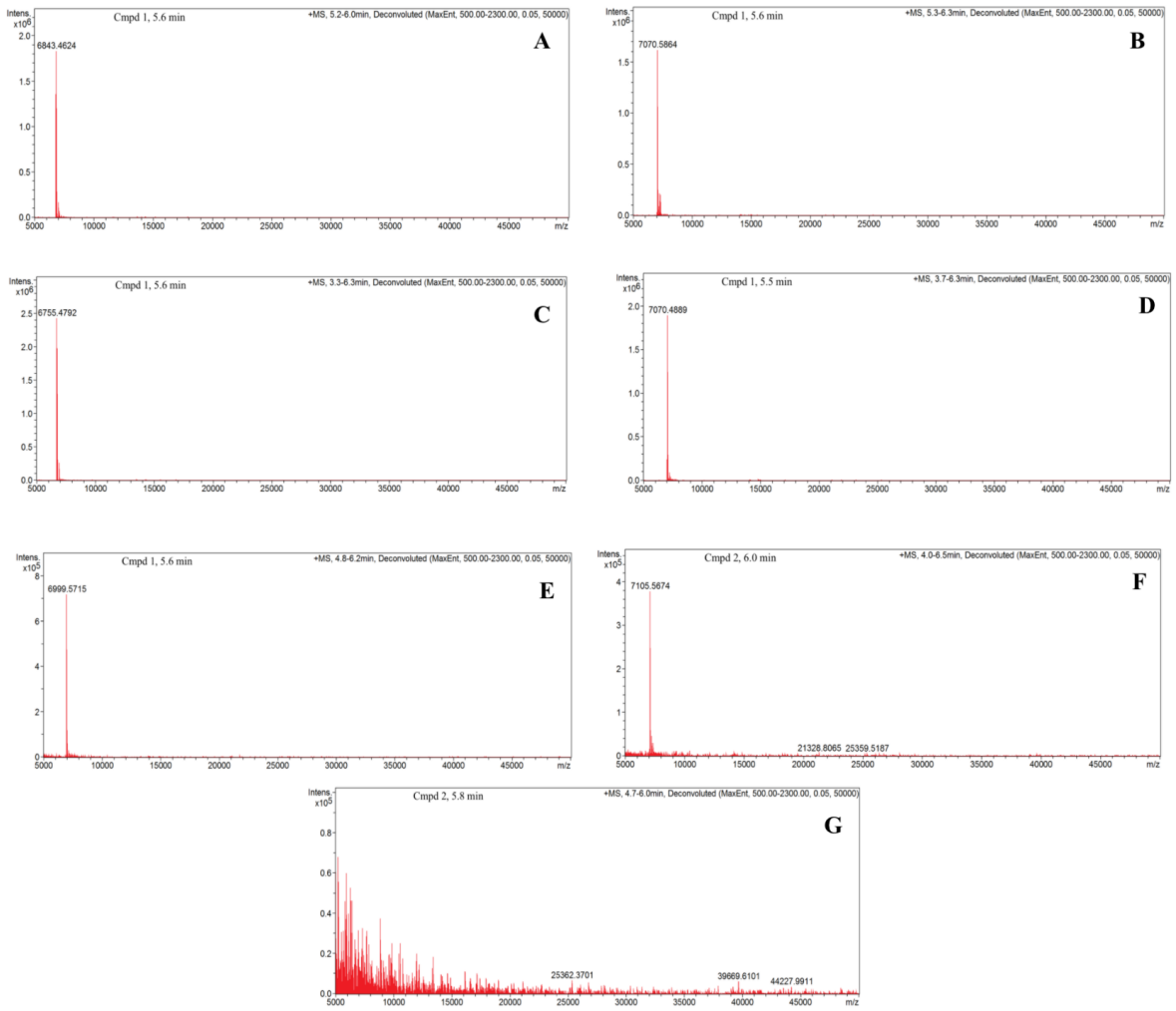



*Figure S3.* ***MS results of ADAPT$_{VEGF}$ variants.*** *Shows the LC/MS chromatograms of ADAPT$_{VEGF-1}$ (A), ADAPT$_{VEGF-2}$ (B), ADAPT$_{VEGF-4}$ (C), ADAPT$_{VEGF-6}$ (D), ADAPT$_{VEGF-7}$ (E), ADAPT$_{VEGF-23}$ (F), ADAPT$_{VEGF-42}$ (G).*

## CD results of ADAPTs

*Table S4.* ***CD results summary.***

| ADAPT | Alpha helical structure at 20°C | Melting temp. (°C) | Refolding after heat denaturation |
|---|---|---|---|
| $ADAPT_{CCL7\_01}$ | Yes | 47 | No |
| $ADAPT_{CCL7\_02}$ | Yes | 35 | No |
| $ADAPT_{CCL7\_03}$ | Yes | 32 | Yes |
| $ADAPT_{CCL7\_05}$ | No | <30 | N/A |
| $ADAPT_{CCL7\_18}$ | Yes | 58 | No |
| $ADAPT_{CCL7\_20}$ | Yes | 44 | Yes |
| $ADAPT_{CCL7\_21}$ | Yes | 42 | Yes |
| $ADAPT_{CCL7\_91}$ | Yes | 62 | Yes |
| $ADAPT_{CEACAM5\_01}$[A] | Yes | 37 | Yes |
| $ADAPT_{CEACAM5\_02}$[A,B] | Yes | 36 | Yes |
| $ADAPT_{CEACAM5\_03}$[A] | Yes | 33 | Yes |
| $ADAPT_{CEACAM5\_04}$[A] | Yes | 72 | Yes |
| $ADAPT_{CEACAM5\_08}$[A] | Yes | 44 | Yes |
| $ADAPT_{CEACAM5\_22}$[A] | No | <30 | N/A |
| $ADAPT_{VEGF\_01}$ | Yes | 50 | Yes |
| $ADAPT_{VEGF\_02}$ | Yes | 52 | Yes |
| $ADAPT_{VEGF\_04}$ | Yes | 48.8 | Yes |
| $ADAPT_{VEGF\_06}$ | Yes | 54 | No |
| $ADAPT_{VEGF\_07}$ | Yes | 49 | Yes |
| $ADAPT_{VEGF\_23}$ | Yes | <30 | No |

[A]VTM program 4 °C – 90 °C. [B]Ran VTM twice. N/A Not applicable.

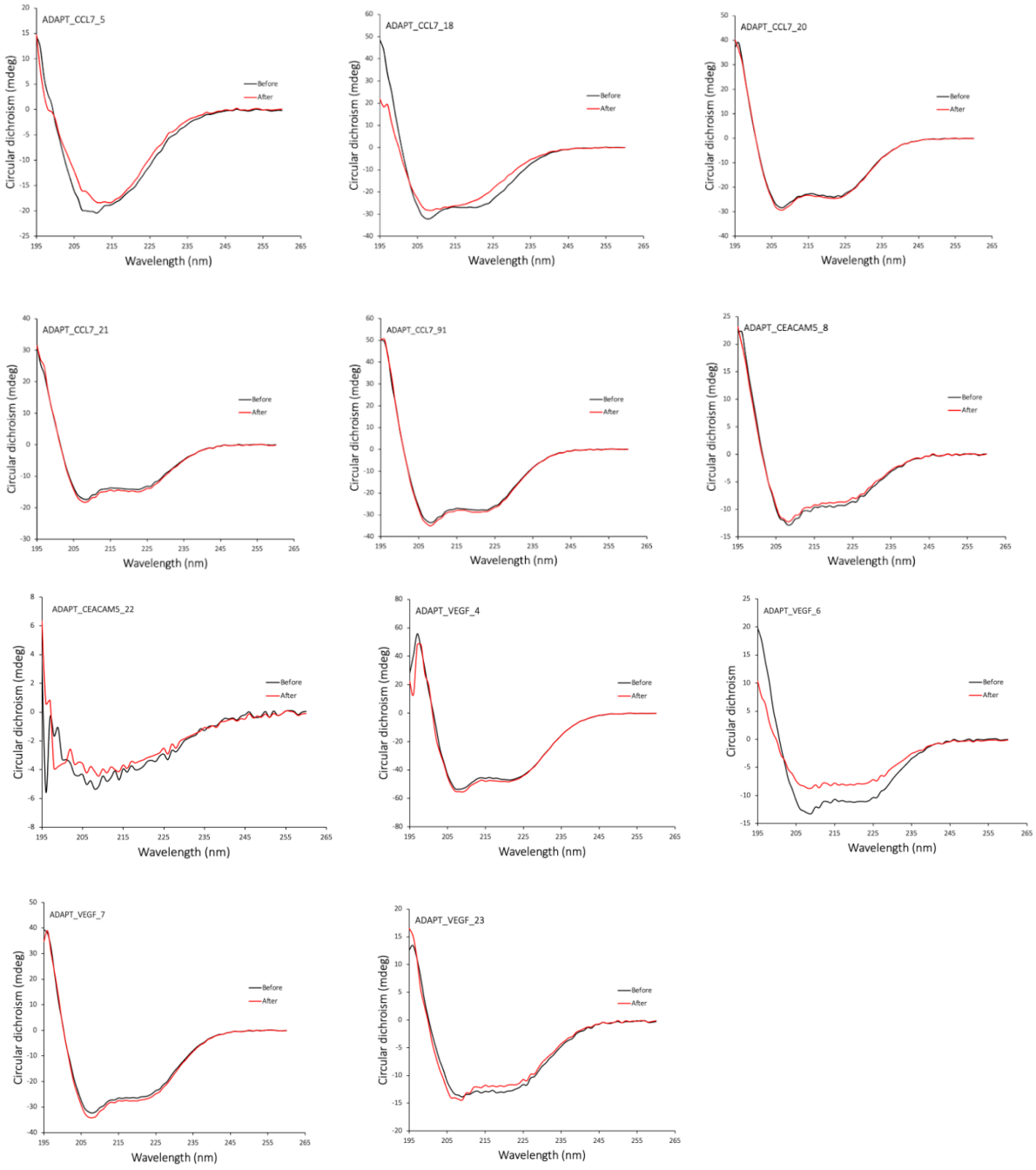



*Figure S4.* ***CD results of ADAPT variants.*** *Shows the CD spectrum of additional* $ADAPT_{CCL7}$, $ADAPT_{CEACAM5}$, $ADAPT_{VEGF}$ *variants that were not included in further analysis.*

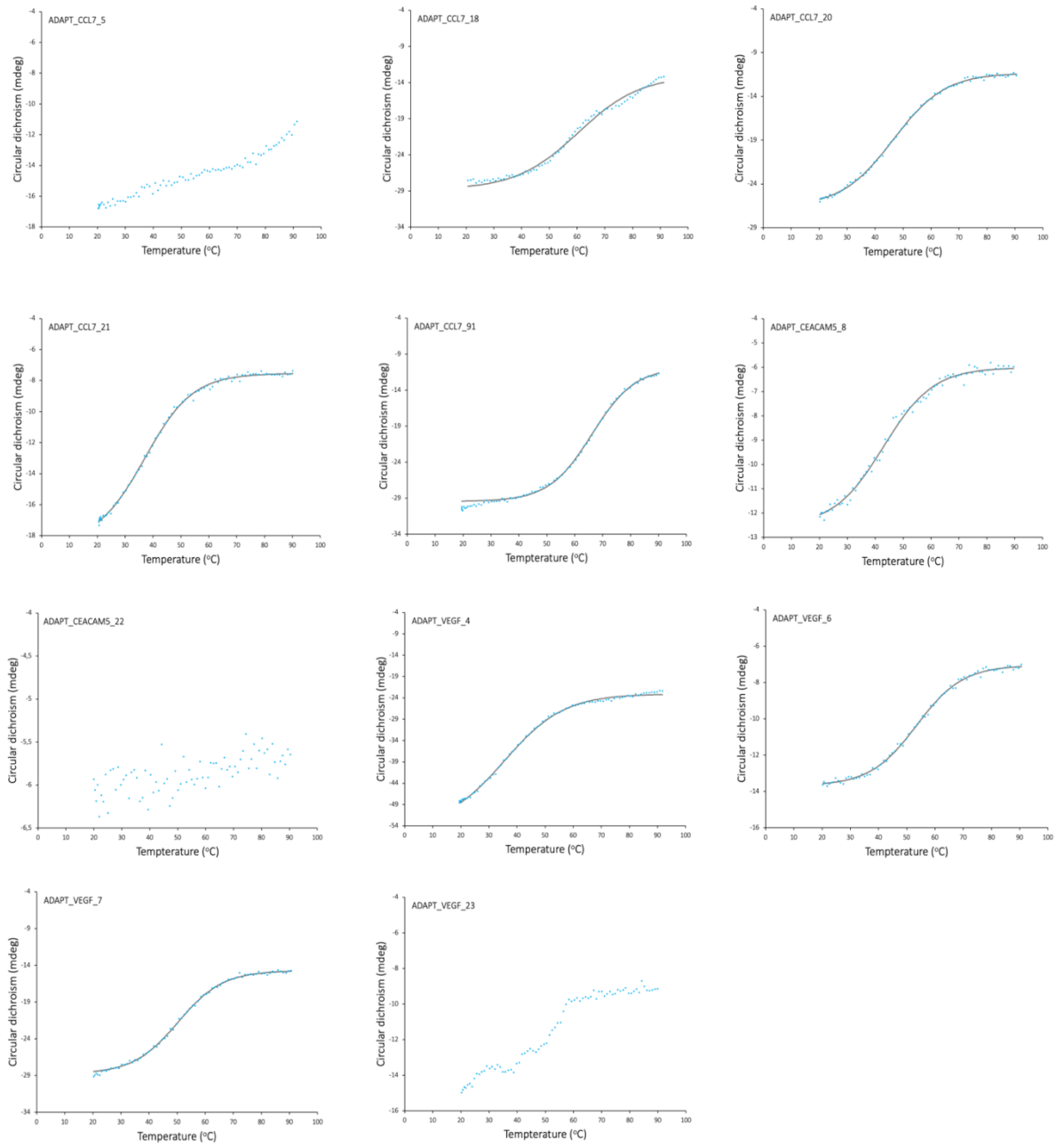



*Figure S5.* ***Melting curves of ADAPT variants.*** *Shows graphs with circular dichroism (mdeg) against temperature (C°) where the measured data points are represented as blue dots and the fitted sigmoid curve as a dark grey line. The melting temperature was determined from these curves, as the temperature at the maximum increase of the circular dichroism.*

## Kinetic values – Simultaneous ADAPTs

*Table S5.* ***Kinetic data of most promising ADAPT variants in SPR.*** *Shows the $K_D$, on-rate ($k_a$) and off-rate ($k_d$) of $ADAPT_{CCL7\text{-}1}$, $ADAPT_{CCL7\text{-}2}$, $ADAPT_{CCL7\text{-}3}$ and $ADAPT_{CEACAM5\_02}$. The kinetic values are calculated from the 1:1 simulating binding between ligand and analyte fitted curves in figure 6.*

| ADAPT | $K_D$ (M) | $k_a$ ($M^{-1}s^{-1}$) | $k_d$ ($s^{-1}$) |
|---|---|---|---|
| $ADAPT_{CCL7\_1}$ | $3.88 \cdot 10^{-8}$ | $9.02 \cdot 10^{8}$ | 35 |
| $ADAPT_{CCL7\_2}$ | $6.21 \cdot 10^{-6}$ | $9.06 \cdot 10^{7}$ | 5.62 |
| $ADAPT_{CCL7\_3}$ | $1.4 \cdot 10^{-7}$ | $8.98 \cdot 10^{7}$ | 13 |
| $ADAPT_{CEACAM5\_2}$ | $3.64 \cdot 10^{-7}$ | $1.29 \cdot 10^{3}$ | $4.7 \cdot 10^{-4}$ |

## Alanine scan - Production & purification data

*Table S6.* ***Production information ADAPT$_{CEACAM5_02}$ mutants.***

| Mutant name | Yield (mg) | Theoretical Mw (Da) | Isoelectric point | Extinction coefficient | MS Determined Mw (Da) |
|---|---|---|---|---|---|
| M1 | 2.53 | 7069.96 | 6.38 | 7450 | 7069.6023 |
| M2 | 1.49 | 7093.94 | 6.45 | 7450 | 7093.5906 |
| M3 | 2.10 | 7078.97 | 6.45 | 7450 | 7077.6042 |
| M4 | 1.52 | 7078.92 | 6.21 | 7450 | 7078.5736 |
| M5 | 0.94 | 7092.01 | 6.76 | 7450 | 7091.6428 |
| M6 | 1.33 | 7043.92 | 6.45 | 5960 | 7043.5995 |
| M7 | 2.07 | 7043.92 | 6.45 | 5960 | 7043.5960 |
| M8 | 0.34 | 7043.92 | 6.45 | 5960 | ND |
| M9 | 1.22 | 7050.91 | 6.21 | 7450 | 7043.6055 |
| M10 | 1.38 | 7092.99 | 6.45 | 7450 | 7091.6229 |
| M11 | 0.87 | 7093.94 | 6.45 | 7450 | 7093.5782 |
| M12 | 0.96 | 7160 | 6.51 | 7450 | 7159.6045 |
| M13 | 0.48 | 7139.08 | 6.45 | 7450 | 7138.6034 |
| M14 | 0.65 | 7109.98 | 6.51 | 5960 | 7109.6200 |
| ADAPT$_{CEACAM5_01}$ | 1.51 | 7137.03 | 6.57 | 5960 | ND |
| ADAPT$_{CEACAM5_02}$ | 1.23 | 7136.02 | 6.45 | 7450 | ND |

ND – not determined.

## Sequences of ADAPT$_{CEACAM5_02}$ based mutants

*Table S7.* ***Amino acid sequences of mutants.*** *The table shows the construction of the alanine mutants. Amino acids in the randomized positions of ADAPT$_{CEACAM5_02}$ are indicated as "X" and are coloured green and amino acids in the unique randomized positions of ADAPT$_{CEACAM5_01}$ are indicated as "Z" and are coloured yellow. The red coloured residues are alanine substitutions.*

| Variant name | Protein sequence |
|---|---|
| ADAPT$_{CEACAM5_02}$ | LXXAKXXAXXELXXYGVSDFYXXLIXKAKTVEGVEALKLHILAALP |
| ADAPT$_{CEACAM5_01}$ | LX ZAK ZXAXXELXZYGVSDFYXXLIXKAKTVEGVEALKLHILAALP |
| M1 | LAXAKXXAXXELXXYGVSDFYXXLIXKAKTVEGVEALKLHILAALP |
| M2 | LXAAKXXAXXELXXYGVSDFYXXLIXKAKTVEGVEALKLHILAALP |
| M3 | LXXAKAXAXXELXXYGVSDFYXXLIXKAKTVEGVEALKLHILAALP |
| M4 | LXXAKXAAXXELXXYGVSDFYXXLIXKAKTVEGVEALKLHILAALP |
| M5 | LXXAKXXAAXELXXYGVSDFYXXLIXKAKTVEGVEALKLHILAALP |
| M6 | LXXAKXXAXAELXXYGVSDFYXXLIXKAKTVEGVEALKLHILAALP |
| M7 | LXXAKXXAXXELAXYGVSDFYXXLIXKAKTVEGVEALKLHILAALP |
| M8 | LXXAKXXAXXELXAYGVSDFYXXLIXKAKTVEGVEALKLHILAALP |
| M9 | LXXAKXXAXXELXXYGVSDFYAXLIXKAKTVEGVEALKLHILAALP |
| M10 | LXXAKXXAXXELXXYGVSDFYXALIXKAKTVEGVEALKLHILAALP |
| M11 | LXXAKXXAXXELXXYGVSDFYXXLIAKAKTVEGVEALKLHILAALP |
| M12 | LXZAKXXAXXELXXYGVSDFYXXLIXKAKTVEGVEALKLHILAALP |
| M13 | LXXAKZXAXXELXXYGVSDFYXXLIXKAKTVEGVEALKLHILAALP |
| M14 | LXXAKXXAXXELXZYGVSDFYXXLIXKAKTVEGVEALKLHILAALP |

## Alanine scan - MS data

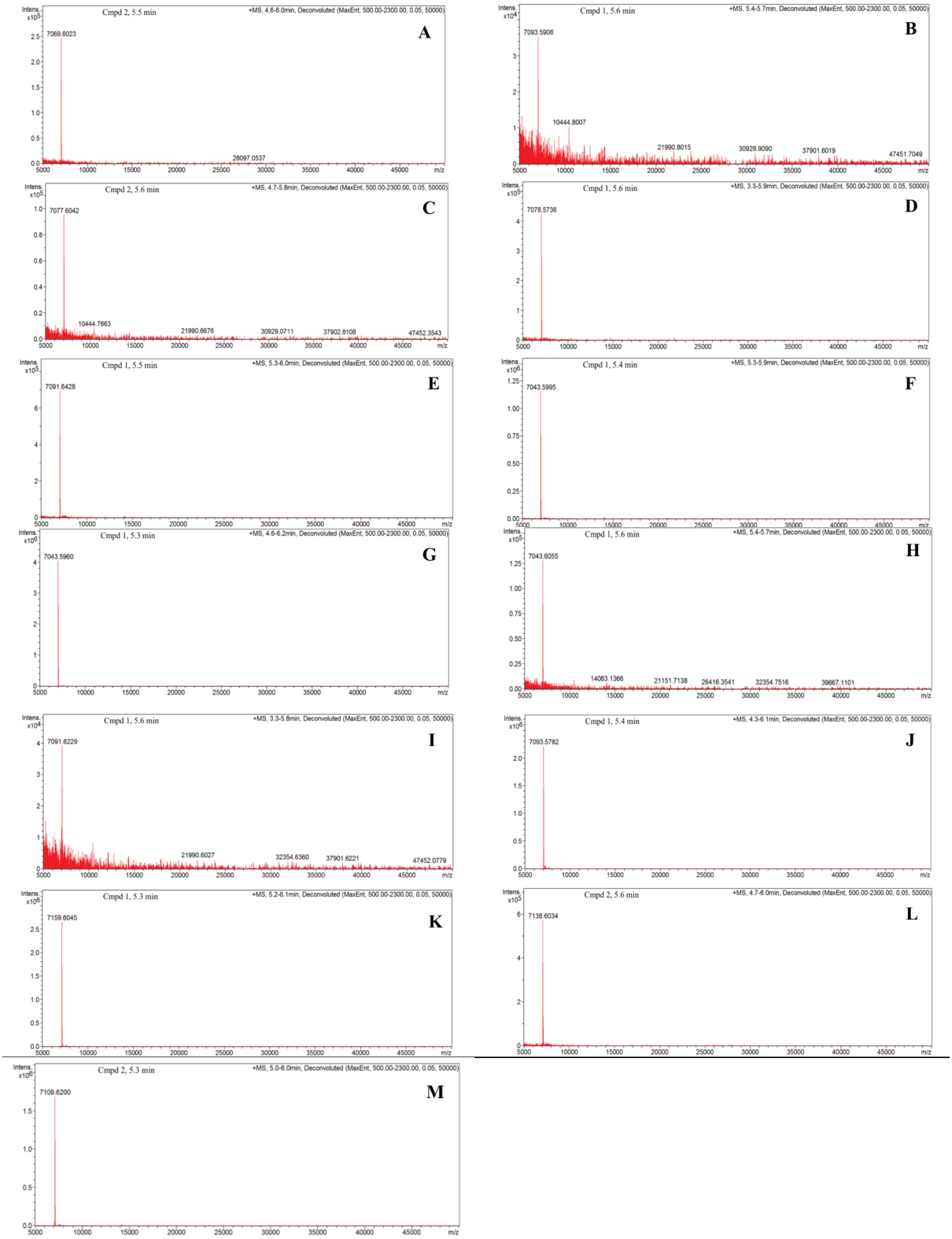



*Figure S6.* ***MS results of ADAPT$_{CEACAM5}$ mutants.*** *Shows the LC/MS chromatogram of M1 (A), M2 (B), M3 (C) , M4 (D), M5 (E), M6 (F), M7 (G), M9 (H), M10 (I), M11 (J), M12 (K), M13 (L), M14 (M).*

## Alanine scan – Supplementary CD graphs

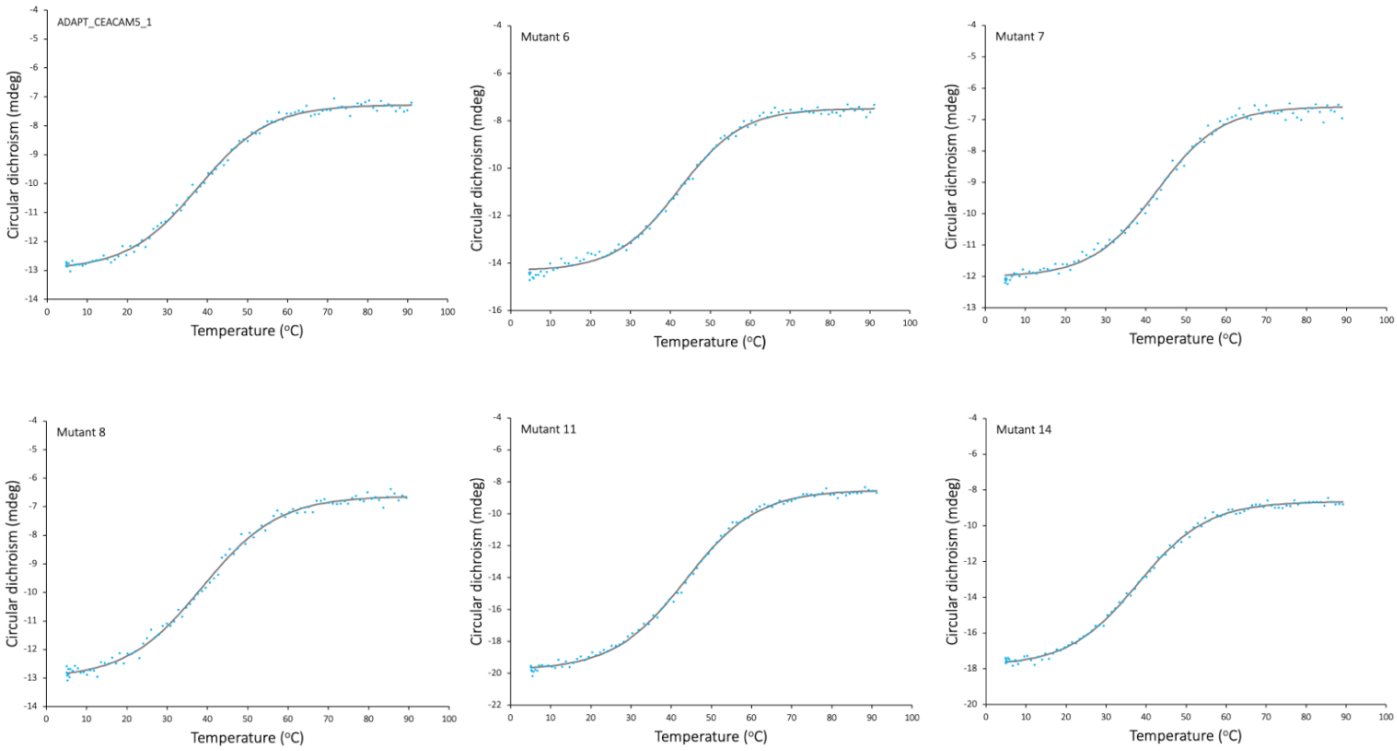



*Figure S7.* ***Melting curves of stable ADAPT$_{CEACAM5}$ mutants and ADAPT$_{CEACAM5_01}$.*** *Shows graphs with circular dichroism (mdeg) against temperature (C°) where the measured data points are represented as blue dots and the fitted sigmoid curve as a dark grey line. The melting temperature was determined from these curves, as the temperature at the maximum increase of the circular dichroism.*

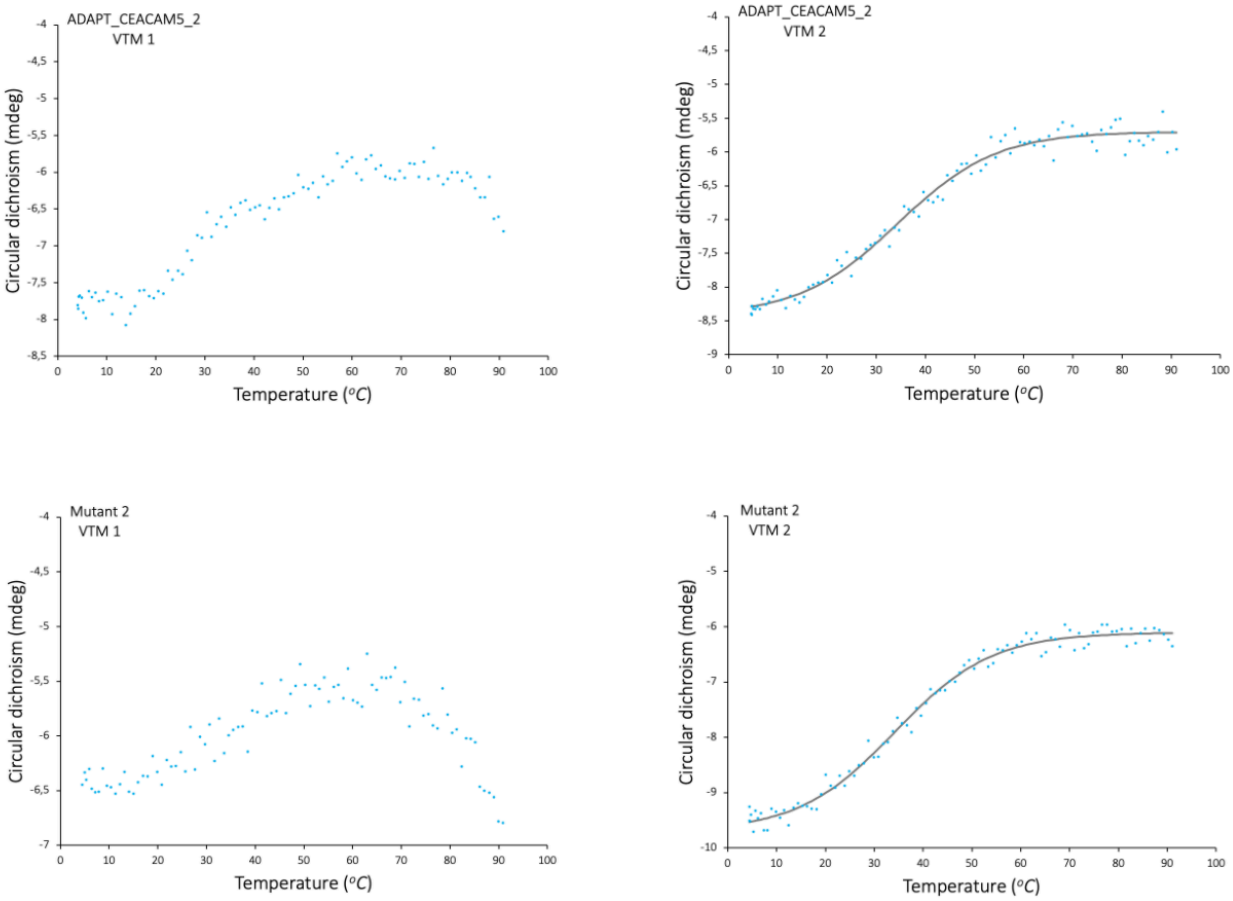



*Figure S8.* ***Melting curves of stable ADAPT$_{CEACAM5_02}$ and mutant 2.*** *Shows graphs with circular dichroism (mdeg) against temperature (C$^{o}$) where the measured data points are represented as blue dots and the fitted sigmoid curve as a dark grey line. The melting temperature was determined from these curves, as the temperature at the maximum increase of the circular dichroism. Here, the VTM was run twice, and clearer melting curves were achieved after second heat treatment (right panels).*

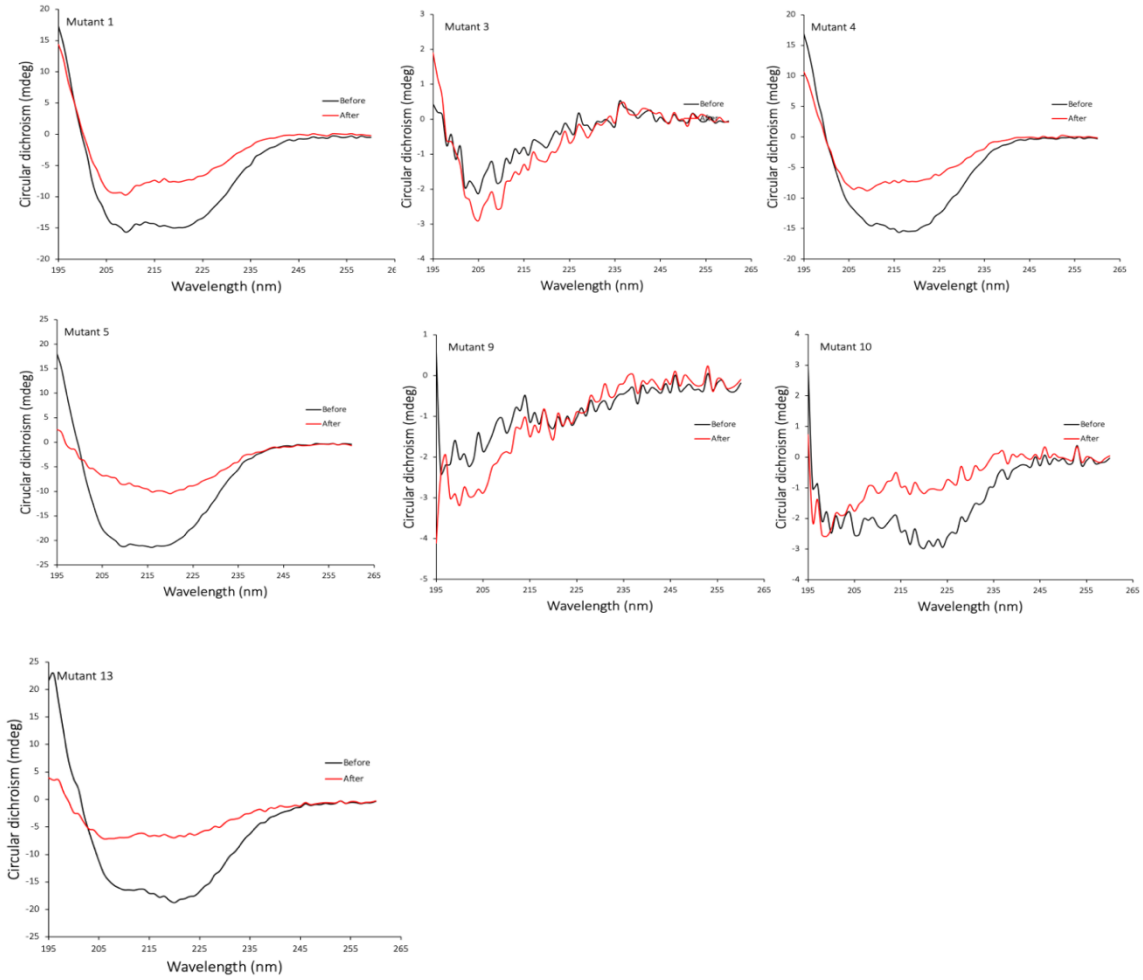



*Figure S9.* ***CD results of unstable ADAPT$_{CEACAM5_02}$ mutants.*** *Shows the mutants without alpha helical structure at 20 °C.*

## Alanine scan – Kinetic values of mutants

*Table S8.* ***Kinetic values of mutants.*** *Shows the on-rate, $k_a$, off-rate, $k_d$, and the dissociation equilibrium constant, $K_D$ calculated from the 1:1 kinetic curve in figure 17 which were fitted to sensorgrams from the T200 SPR instrument.*

| ADAPT | $k_a$ ($M^{-1}s^{-1}$) | $k_d$ ($s^{-1}$) | $K_D$ value CEACAM5 (M) |
|---|---|---|---|
| M1 | $5.03 \cdot 10^{3}$ | $4.27 \cdot 10^{-4}$ | $4{,}48 \cdot 10^{-8}$ |
| M2 | $6.3 \cdot 10^{3}$ | $2.41 \cdot 10^{-4}$ | $3.83 \cdot 10^{-8}$ |
| M3 | $5.85 \cdot 10^{3}$ | $1.80 \cdot 10^{-4}$ | $3.08 \cdot 10^{-8}$ |
| M4 | $5{,}35 \cdot 10^{3}$ | $1.43 \cdot 10^{-4}$ | $2.67 \cdot 10^{-8}$ |
| M5 [a] | 523 | 0.0116 | $2.22 \cdot 10^{-5}$ |
| M10 | $4.31 \cdot 10^{3}$ | $9.17 \cdot 10^{-4}$ | $2.13 \cdot 10^{-7}$ |
| M12[b] | $4.98 \cdot 10^{2}$ | $1.52 \cdot 10^{-4}$ | $3.05 \cdot 10^{-7}$ |
| M13[a] | $7.74 \cdot 10^{3}$ | $5.15 \cdot 10^{-5}$ | $6.65 \cdot 10^{-9}$ |
| M14[a] | $6.63 \cdot 10^{3}$ | $3.07 \cdot 10^{-4}$ | $4.63 \cdot 10^{-8}$ |
| $ADAPT_{CEACAM5\_02}$ | $7.6 \cdot 10^{3}$ | $1.2 \cdot 10^{-4}$ | $1.58 \cdot 10^{-8}$ |

*[a]Poor fits, kinetic values are only an estimation. [b]Values and graphs made in the 8K instrument.*