



HAL
open science

Criblage phénotypique à l'aide d'intracorps dans un modèle de cancer colorectal

Vincent Parez

► **To cite this version:**

Vincent Parez. Criblage phénotypique à l'aide d'intracorps dans un modèle de cancer colorectal. Biologie cellulaire. Université Montpellier II - Sciences et Techniques du Languedoc, 2014. Français. NNT : 2014MON20109 . tel-01528641

HAL Id: tel-01528641

<https://theses.hal.science/tel-01528641>

Submitted on 29 May 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de
Docteur

Délivré par **UNIVERSITE MONTPELLIER 2**

Préparée au sein de l'école doctorale CBS2
et de l'unité de recherche IRCM – INSERM U896

Spécialité : **Biologie Cellulaire et Moléculaire**

Présentée par **Vincent PAREZ**

**CRIBLAGE PHENOTYPIQUE A L'AIDE D'INTRACORPS
DANS UN MODELE DE CANCER COLORECTAL.**

Soutenue le 30/10/2014 devant le jury composé de

Dr Gisèle CLOFENT-SANCHEZ, DR, CNRS
Dr Etienne WEISS, PR, Université Strasbourg
Dr Pierre MARTINEAU, CR, INSERM
Dr Paul GUGLIELMI, DR, INSERM
Dr Piona DARIAVACH, MCU, Université Montpellier 2

Rapporteur
Rapporteur
Examineur
Invité
Directrice de thèse

THÈSE

Pour obtenir le grade de
Docteur

Delivered by the **UNIVERSITY MONTPELLIER 2**

Prepared at the doctoral school CBS2
and the research unit IRCM – INSERM U896

Specialization: **Molecular and Cellular Biology**

Presented by **Vincent PAREZ**

**A PHENOTYPIC SCREEN USING INTRABODIES IN A
COLORECTAL CANCER MODEL.**

Defended on 30/10/2014 in front of a jury composed of

Dr Gisèle CLOFENT-SANCHEZ, DR, CNRS
Dr Etienne WEISS, PR, Université Strasbourg
Dr Pierre MARTINEAU, CR, INSERM
Dr Paul GUGLIELMI, DR, INSERM
Dr Piona DARIAVACH, MCU, Université Montpellier 2

Rapporteur
Rapporteur
Examineur
Invité
Directrice de thèse

Mots clés : Criblage phénotypique, anticorps intracellulaires, cancer colorectal

Criblage phénotypique à l'aide d'intracorps dans un modèle de cancer colorectal.

L'expression intracellulaire des anticorps (intracorps) est une approche qui permet l'étude et le ciblage des antigènes dans les compartiments intracellulaires. Néanmoins, l'expression d'anticorps entiers fonctionnels dans les cellules reste une tâche difficile en raison de leur grande taille et de leur structure, l'environnement réducteur du milieu intracellulaire étant défavorable à la formation des ponts disulfure. Notre groupe a une forte expertise dans le domaine de l'immunisation intracellulaire et son application pour l'identification de nouvelles cibles thérapeutiques. Pour cela, notre équipe a élaboré des banques de fragments d'anticorps scFv optimisés pour une meilleure expression intracellulaire. Nos travaux antérieurs ont démontré que ces intracorps peuvent cibler spécifiquement des domaines ou des modifications post-traductionnelles de protéines dans des cellules vivantes. Ceci est particulièrement important car il démontre l'un des avantages principaux des intracorps par rapport à l'approche basée sur l'ARNi. Cet avantage a été démontré par un criblage phénotypique dans un modèle d'allergie. En appliquant cette approche à l'étude de l'activation des mastocytes, nous avons pu identifier un nouvel acteur moléculaire impliqué dans la voie de signalisation mise en jeu. Ce travail a été protégé par un brevet européen en 2013 et est publié récemment. Dans le cadre de mon projet de thèse, j'ai construit une nouvelle banque synthétique (HUSCiv) optimisée pour la stabilité, la diversité et l'affinité des scFvs. Pour cela, le scFv 13R4 isolé dans notre équipe a servi de charpente pour le greffage des différentes boucles hypervariables, tout en respectant la diversité des régions CDR observée dans les anticorps naturels humains. Nous avons utilisé la protéine GFP en tant que rapporteur pour étudier le repliement et la solubilité des intracorps. Nos résultats ont clairement démontré que la plupart des intracorps issus de la banque HUSCiv sont solubles dans le cytoplasme des cellules mammifères.

Mon projet de thèse décrit ici rapporte l'utilisation de la banque HUSCiv pour un criblage phénotypique dans des cellules de cancer colorectal portant une mutation du gène *KRAS* et résistantes au traitement par l'anticorps chimérique Cetuximab. Le projet cherche à sélectionner des

scFv capables de restaurer la sensibilité au Cetuximab, avec comme objectif l'identification des cibles intracellulaires impliquées. Pour ce criblage fonctionnel, la banque HUSCIV a été exprimée dans les cellules HCT116 par l'intermédiaire d'un système d'expression rétroviral. Le processus de sélection est basé sur la sélection directe de la prolifération des cellules en utilisant un colorant fluorescent (CMRA). Les cellules dont la prolifération est bloquée sont isolées et un séquençage à haut débit permet de suivre l'évolution des populations de scFv tout au long de l'expérience. Ainsi, ce projet a nécessité un séquençage profond d'un grand nombre de scFv afin de réaliser une analyse statistique.

Nous avons réalisé à ce jour deux tours de sélection. Les tests de cytotoxicité réalisés sur les populations sélectionnées ont montré une inhibition significative de la prolifération en présence du Cetuximab d'environ 10%. Ces résultats indiquent l'évolution du phénotype qui tend vers une sélection de scFv inhibiteurs et suggèrent que nous devons réaliser au moins un ou plusieurs tours plus sélectifs avant de formuler des conclusions.

L'approche introduite ici est différente de toutes les études existantes en ce qu'elle utilise des banques « naïves », et permet non seulement de répondre à la diversité du protéome, mais aussi d'étudier les messagers secondaires et le métabolisme des cellules. En tant que tel, et par rapport à d'autres approches à grande échelle, celle-ci représente une voie simple pour la découverte de molécules thérapeutiques potentielles.

Keywords: Phenotypic screen, intracellular antibodies, colorectal cancer

A phenotypic screen using intrabodies in a colorectal cancer model.

Intracellular expression of antibodies (intrabodies) permitted the study and targeting of antigens in cellular compartments. However, the expression of functional intrabodies remains a difficult task due to their large size, structure, and the reducing intracellular environment. Our group has a strong expertise in the field of intracellular immunization and the identification of new therapeutic targets. For this purpose, we have developed an scFv library optimized for intracellular expression of scFv intrabody fragments. Our previous works have shown the successful use of intrabodies for targeting specific domains or post-translational modifications in living cells. This is particularly important because it demonstrates one of the main advantages of intrabodies compared to the approaches using RNAi. This benefit was demonstrated by a phenotypic screen in a model of allergy. Applying this approach to the study of mast cell activation, we identified a new molecular player involved in the signaling pathway implemented. This work was protected by a European patent in 2013 and was recently published. As part of my thesis project, I designed a new synthetic library (HUSCiv) optimized for scFv stability, diversity and affinity. For this, a highly soluble and hyper-stable framework, scFv13R4 isolated in our group, was used as a scaffold for grafting different hypervariable loops, while respecting the diversity of CDRs observed in human natural antibodies. We used protein GFP as a reporter to study the folding and solubility of intrabodies. Our findings clearly demonstrated that most of the intrabodies from HUSCiv library are soluble in the cytoplasm of mammalian cells.

My thesis project described here reports the use of HUSCiv in a phenotypic screen of colorectal cancer cells carrying a mutation in the *KRAS* gene and resistant to the treatment with the chimeric antibody Cetuximab. The project seeks to select scFv fragments able to restore the sensitivity to Cetuximab, with the objective to identify the intracellular targets involved. For this functional screen, the HUSCiv library was expressed in HCT116 cells via a retroviral expression system. The selection process is based on the direct selection of cell proliferation using a fluorescent dye (CMRA). The cells whose proliferation is blocked are isolated and the evolution of scFv populations throughout the experiment is tracked via high-throughput sequencing. This sequencing requires a large number of scFvs to perform a statistical analysis.

So far, we have achieved two rounds of selection. The cytotoxicity tests carried out on the selected populations showed a significant inhibition of proliferation (10%) in the presence of Cetuximab. These results indicate that the evolving phenotypes are tending towards a selection of scFv inhibitors and suggest that we need to perform at least one or more selective rounds before making conclusions.

The approach introduced here is different from all existing studies in that it uses "naive" libraries not only to respond to the diversity of the proteome, but also to study secondary messengers and metabolism in cells. As such, and in comparison to other large-scale approaches, it is a simple way for the discovery of potential therapeutic molecules.

1. Introduction

L'approche de « l'immunisation intracellulaire » est basée sur la spécificité de reconnaissance et la diversité du répertoire des anticorps, permettant de cibler, neutraliser ou moduler les fonctions des protéines intracellulaires. Les anticorps intracellulaires sont alors appelés « intracorps ». L'expression intracellulaire des anticorps remonte aux années 80, et ces premières expériences ont permis l'utilisation des anticorps comme outils pour l'étude et le ciblage des protéines intracellulaires. Néanmoins, l'expression d'anticorps entiers fonctionnels dans les cellules reste une tâche difficile en raison de leur grande taille et de leur structure, l'environnement réducteur du milieu intracellulaire étant défavorable à la formation des ponts disulfure. Aussi, l'ingénierie des anticorps a permis de développer des formats de plus petite taille capables de retenir les mêmes propriétés d'affinité et de spécificité pour un antigène que les anticorps parents, et plus adaptés pour une utilisation intracellulaire. Des vecteurs dédiés à l'expression d'intracorps ont alors été développés. Ces vecteurs sont disponibles sous différentes versions, permettant l'expression des fragments d'anticorps dans les divers compartiments subcellulaires comme le cytoplasme, le noyau, ou le réticulum endoplasmique. En plus, le développement de la biologie moléculaire notamment grâce à la PCR et au séquençage de la totalité des gènes codant pour les immunoglobulines, a facilité la manipulation et le sous-clonage des gènes codant les anticorps.

Le format scFv correspond au plus petit fragment conservant l'ensemble du site de liaison à l'antigène. Il est constitué d'une séquence nucléotidique unique codant les domaines variables des chaînes lourdes et légères (VH et VL), reliés entre eux par un peptide de liaison flexible. En raison de leur petite taille les scFv offrent de nombreux avantages, d'autant plus qu'ils peuvent conserver la même affinité et spécificité de liaison élevée pour un antigène qu'un anticorps entier. La principale limite de ce format est que de nombreux scFv manquent de stabilité en raison de l'absence de ponts disulfure intra-domaines, et se retrouvent sous forme d'agrégats insolubles. Cependant des études ont montré que dans ces conditions, et malgré leur agrégation, certains intracorps conservaient leurs propriétés d'affinité de liaison et étaient capables de séquestrer leur cible antigénique. La composition en acides aminés des scFv semble être déterminante pour leur stabilité intracellulaire.

Les formats d'intracorps les plus utilisés sont les scFv, ainsi que le format simple domaine qui commence à émerger pour de telles applications. Le choix du format dépendra de l'application

souhaitée, chacun présentant des différences en termes de taille, d'affinité et de stabilité. Enfin, un certain nombre de banques synthétiques d'intracops ont été élaborées et ont permis de générer des fragments d'anticorps optimisés pour des applications intracellulaires.

La facilité de production de fragments d'anticorps fonctionnels en grande quantité chez *E. Coli* a conduit des équipes à construire des banques d'anticorps recombinants, ainsi qu'à réaliser des criblages de telles banques dans le but d'isoler les anticorps spécifiques d'un antigène donné, ou encore exprimés de manière stable dans le milieu cytosolique réducteur des cellules. Généralement, la probabilité d'obtenir un anticorps recombinant spécifique pour un antigène donné augmente proportionnellement avec la taille de la banque utilisée. Au sein de notre équipe, une banque synthétique (PMEW) optimisée pour l'expression intracellulaire a été construite en utilisant la séquence du scFv13R4 comme charpente, la diversité ayant été introduite au niveau des régions CDR3. Des travaux antérieurs avaient permis d'optimiser ce scFv par évolution moléculaire dirigée pour une bonne expression et activité cytoplasmique. L'inconvénient majeur des anticorps d'origine murine ou humaine est leur manque de stabilité et leur tendance à l'agrégation. C'est pourquoi une nouvelle banque d'intracops humain (HUSCI) a été conçue afin d'améliorer la stabilité, la diversité et l'affinité des fragments d'anticorps (chapitre 3).

De manière concomitante à l'élaboration des banques d'anticorps recombinants, plusieurs techniques se sont développées pour la sélection des fragments d'anticorps contre des ligands donnés. Le principe de ces méthodes de sélection repose sur l'association entre un phénotype et le génotype. La technique de *Intracellular Ab Capture* couple deux méthodes. Cette technique permet à la fois de sélectionner des scFv dirigés contre une cible précise, et de confirmer cette interaction à l'intérieur de cellules eucaryotes. Une fois exprimés les intracops peuvent se lier à leur cible antigénique intracellulaire. Cette interaction peut résulter en la modulation ou l'inhibition des fonctions de l'antigène soit par une interférence directe, soit par une délocalisation subcellulaire. En effet, la liaison d'un intracops à sa cible peut empêcher des interactions protéine-protéine ou protéine-acide nucléique, mais il est également possible d'élaborer des intracops contenant des signaux d'adressage vers les divers compartiments subcellulaires pour délocaliser ou séquestrer une cible protéique.

De cette façon des propriétés spécifiques de protéines peuvent être exploitées ou détournées par l'utilisation d'intracorp, pour induire ou moduler un phénotype associé à la protéine cible. On peut définir cette utilisation comme de la mutagenèse à l'échelle de la protéine. Notre groupe a une forte expertise dans le domaine de l'immunisation intracellulaire et son application pour l'identification de nouvelles cibles thérapeutiques et de nouveaux médicaments. Des travaux antérieurs effectués dans notre groupe ont démontré l'efficacité de cette approche par l'inhibition des interactions protéine-protéine et par la découverte de nouvelles molécules thérapeutiques. Le laboratoire a également démontré que les anticorps intracellulaires peuvent cibler spécifiquement une modification post-traductionnelle d'une protéine dans une cellule vivante. Ceci est particulièrement important car il démontre l'un des avantages principaux des intracorp par rapport à l'approche basée sur l'ARNi.

Dans le chapitre suivant, je décrirai les travaux que j'ai effectués au début de ma thèse de Doctorat dans le cadre d'un criblage phénotypique dans un modèle d'allergie qui a conduit à l'identification d'une nouvelle cible thérapeutique impliquée dans l'activation des mastocytes. Ce travail a été protégé par un brevet européen en 2013 et vient d'être publié. Le chapitre 3 de cette introduction résume mon travail pour l'optimisation de la banque de fragments scFv (HUSCI) afin d'améliorer leur expression et leur stabilité intracellulaire dans des conditions réductrices. Pour terminer dans le chapitre 4, je décrirai l'utilisation de la banque HUSCIV optimisée pour un criblage phénotypique dans des cellules de cancer colorectal portant un gène *KRAS* muté et résistantes au traitement par l'anticorps chimérique Cetuximab. Mon objectif était d'identifier des anticorps capables de rétablir la sensibilité au traitement, et par la suite d'identifier leurs cibles cellulaires.

2. Criblage phénotypique dans un modèle d'allergie

Les mastocytes dérivent de progénitures de la moelle osseuse qui entament leur maturation pendant leur déplacement à travers les tissus périphériques et la terminent localement une fois le site où ils vont résider atteint. Ils sont présents dans tous les organes et tissus, mais plus particulièrement associés avec le tissu conjonctif, ainsi qu'aux interfaces avec le milieu extérieur. Une fois matures, ils expriment le récepteur haute-affinité pour les IgE (FcεRI), et présentent des granules intracellulaires contenant de nombreux médiateurs pro-inflammatoires.

En réponse à une stimulation, ils sont capables de relarguer par exocytose leur contenu dans des granules de sécrétion ou synthétisés, c'est ce que l'on appelle la dégranulation. La voie la mieux

décrite d'activation des mastocytes est celle de la liaison des IgE couplées à leur antigène, aux récepteurs de haute affinité FcεRI, ce qui conduit à l'agrégation des FcεRI et déclenche l'activation de cascades intracellulaires aboutissant à la libération de messagers par exocytose. Ces signaux précoces de fixation des IgE impliquent le rapprochement des récepteurs agrégés au sein des « lipid rafts », et l'activation de Lyn, une tyrosine kinase de la famille Src, qui va phosphoryler les résidus tyrosines présents dans les motifs ITAM du récepteur. Une fois phosphorylés, ces motifs vont permettre le recrutement de Lyn et de la tyrosine kinase Syk par leurs domaines SH2. L'ancrage de Syk va permettre son auto- ainsi que sa transphosphorylation par Lyn, ce qui va l'activer et lui permettre de phosphoryler un certain nombre de substrats, impliqués dans l'activation mastocytaire.

En appliquant la technique de criblage phénotypique à l'étude de l'activation des mastocytes, notre équipe a pu identifier un nouvel acteur moléculaire impliqué dans la voie de signalisation mise en jeu. Notre but était de sélectionner, à partir d'une banque combinatoire de fragments d'anticorps scFv (PMEW) exprimée dans une lignée mastocytaire, ceux capables d'induire une inhibition de la libération de médiateurs allergiques. Pour cela, la banque PMEW d'une grande diversité et optimisée pour une meilleure expression intracellulaire a été clonée dans deux vecteurs d'expression eucaryotes : un vecteur plasmidique et un vecteur rétroviral. Une fois les banques de scFv exprimées dans la lignée mastocytaire RBL-2H3, deux sélections phénotypiques ont été réalisées en parallèle par cytométrie en flux (FACS), afin de sélectionner les cellules présentant un défaut d'activation et exprimant à priori des scFv inhibiteurs. Plusieurs scFv ont ainsi été sélectionnés et analysés. Les scFv isolés au moyen de l'approche plasmidique ont été analysés indépendamment sur des transfectants stables de la même lignée mastocytaire. L'évolution de la diversité de la banque et l'enrichissement en scFv au moyen de la sélection rétrovirale a été également étudiée par séquençage haut-débit. Ces analyses ont révélé qu'une vingtaine de séquences avaient été enrichies, et parmi elles un fragment d'anticorps en particulier, le 5H4 a retenu notre attention car il a été isolé également au moyen de la sélection plasmidique.

Des expériences d'immunoprécipitation couplées à une analyse par spectrométrie de masse, ont permis l'identification de la cible de l'anticorps 5H4. Il s'agit de la protéine LOC297607/C12orf4, dont la fonction était jusqu'alors inconnue. Nous avons confirmé l'implication de cette protéine dans l'activation des mastocytes par une approche utilisant des shRNA.

En résumé, la convergence de l'ensemble de nos résultats, renforcés par l'identification d'un nouvel acteur moléculaire dans le domaine de l'allergie et l'inflammation, démontre le fort potentiel du criblage phénotypique développé durant ce projet. Cette approche offre de nouvelles perspectives d'études fonctionnelles du protéome, avec des applications potentielles à l'étude de multiples phénotypes pathologique.

3. La banque d'intracorps HUSCI: Conception et Synthèse

L'immunisation intracellulaire utilise des fragments d'anticorps exprimés à l'intérieur de la cellule afin de se lier à une protéine et d'inhiber ou de moduler sa fonction. Le problème majeur limitant l'utilisation des intracorps est l'absence de formation des ponts disulfures en milieu cytoplasmique (réducteur). Nous avons conçu une nouvelle banque de scFvs en nous basant sur les résultats obtenus avec la banque PMEW afin d'améliorer les propriétés de stabilité et de diversité des intracorps. Toutefois, devant les difficultés rencontrées pour obtenir des intracorps par greffages des boucles hypervariables, nous avons créé une banque d'anticorps optimisés pour l'immunisation intracellulaire. Pour cela, différentes boucles hypervariables ont été greffées sur la charpente de l'anticorps 13R4 en respectant la diversité des CDR observée dans les anticorps naturels humains. Notre but était de permettre la production d'une majorité de scFvs sous forme soluble et active dans des cellules mammifères, et avec des hauts rendements de purification dans *E. Coli*. Il était également important d'optimiser les intracorps afin d'améliorer leur spécificité pour les modifications conformationnelles et post-traductionnelles. La stratégie consistait à garder la même base hyper-stable du 13R4 en optimisant les positions clés à diversifier. La diversité de la banque PMEW est située sur toutes les positions des CDR3s et exclusivement sur les CDR3s. Dans HUSCI, les positions diversifiées sont distribuées sur les 6 CDRs, et correspondent à des positions clés en termes de liaison à l'antigène et de stabilité. Les critères se trouvent dans la chaîne latérale impliquée dans la liaison à l'antigène et dans la chaîne latérale non-impliquée dans la stabilité du scFv. Enfin, dans la banque PMEW, les longueurs des CDR3s ont été variées en suivant la distribution en longueurs des anticorps humains. Dans HUSCI, par souci de stabilité, seules les longueurs du H3 sont variables: les 8 longueurs les plus courantes sont représentées avec la même fréquence.

Le choix des positions et résidus à diversifier est dépendant de leur contribution énergétique de liaison à l'antigène et de liaison intra-scFv (stabilité). Ces mesures d'énergies par « computational alanine scanning » ont nécessité un travail préalable de construction et de formatage d'une banque de donnée de structures cristallographiques d'anticorps-antigène. Deux informations principales ressortent de cette analyse. Tout d'abord, tous les CDR contribuent significativement à l'énergie d'interaction antigène-anticorps et deuxièmement, l'énergie d'interaction est distribuée de façon inhomogène, à des positions clefs des CDRs. Les CDR3 (uniques CDRs variables dans la précédente banque PMEWS) ne contribuent en moyenne qu'à hauteur de 40% de l'énergie totale de liaison anticorps-antigène dans les structures de la PDB. Même dans le cas de l'unique structure issue de la banque PMEWS, les CDRs variables (H3 et L3) ne contribuent qu'à hauteur de 30% de l'énergie totale. Les résultats de cette analyse confirment le bien-fondé d'une approche ciblant les positions clefs de l'interaction anticorps-antigène sur les 6 CDRs, et permet de spécifier les positions et les types de résidus optimaux à choisir. Cette distribution en acides aminés correspond à la distribution de la banque de données PDB aux positions clefs en termes de propriétés physico-chimique. Chaque position mutée a été validée en termes d'accessibilité sur des structures modèles représentatives pour chacune des 8 longueurs de H3.

Pour évaluer les scFv issus de la nouvelle banque (HUSCIV) en tant qu'intracorps, ils ont été comparés à ceux de l'ancienne banque (PMEWS). Pour cela, les scFv ont été exprimés dans une lignée cellulaire humaine (HeLa), et dans la lignée cellulaire modèle de cancer colorectal (HCT116). Afin d'étudier le repliement des scFv dans le cytoplasme de cellules de mammifères, une fusion à la protéine GFP de méduse a été réalisée. L'utilisation des banques d'intracorps à grande échelle dépend de plusieurs paramètres. Tout d'abord, les scFv doivent être faciles à isoler. Deuxièmement, les scFv devraient être capables de se replier dans tous les compartiments de la cellule, en particulier dans les compartiments réducteurs. La comparaison entre PMEWS et HUSCIV démontre que la nouvelle banque a été améliorée en termes d'expression et de repliement des intracorps. En effet, le marquage GFP utilisé en tant que marqueur a clairement démontré que la plupart des intracorps issus de la banque HUSCIV sont solubles dans le cytoplasme de la cellule. De plus, le niveau d'expression de tous les clones est comparable, ce qui limite le risque d'artefacts lors de la sélection. De plus, alors que la banque PMEWS contient environ 30% de scFv13R4, la banque HUSCIV en contient deux fois moins. Considérant que HUSCIV contient des intracorps plus solubles que PMEWS, nous avons conclu que les mutations introduites dans les six boucles de CDR n'ont pas réduit la solubilité globale de HUSCIV.

4. Criblage phénotypique dans un modèle de cancer colorectal résistant au Cetuximab

A l'Institut du Cancer de Montpellier, plusieurs équipes ont déjà criblé la banque HUSCI et des anticorps contre leurs cibles respectives (p38 MAP kinase, la cathepsine-D, HER4, famille TAM ...) ont été sélectionnés avec succès par phage display. Notre projet, cependant, introduit cette banque dans une approche totalement différente. En effet, nous avons cloné la banque HUSCIv dans un système rétroviral et nous l'avons utilisée dans le cadre d'un criblage phénotypique visant à sélectionner des scFv capables de restaurer la sensibilité au Cetuximab des cellules de cancer colorectal (mCRC) exprimant un gène *KRAS* muté.

Le cancer du côlon est la deuxième ou troisième pathologie cancéreuse (suivant le sexe) avec environ 40 500 cas et 17 500 morts par an en France. Le traitement neoadjuvant actuel des métastases du mCRC s'appuie sur l'association entre une chimiothérapie à base de deux ou trois molécules (5-fluorouracile, oxaliplatine et/ou irinotecan) et une thérapie ciblée à base d'anticorps monoclonal anti-VEGF (Bevacizumab) ou anti-EGFR (Cetuximab). Dans les meilleures études, le taux de réponse objective reste toutefois limité à environ 50-60% des patients et, en l'absence de chirurgie secondaire, 100% des patients rechuteront et deviendront résistants à toute nouvelle chimiothérapie. L'apparition de résistances au traitement a été identifiée comme la principale raison des rechutes et des échecs thérapeutiques. Cette résistance peut être préexistante au traitement, la tumeur étant intrinsèquement résistante, ou apparaître au cours de la chimiothérapie, les cellules résistantes étant sélectionnées et enrichies par le traitement lui-même. De nombreux mécanismes peuvent être responsables de la réponse ou de la non-réponse des patients souffrant de mCRC aux thérapies mais l'un des mieux établis est l'influence des mutations de la voie RAS en aval du récepteur à l'EGF. En effet, les patients dont les tumeurs sont mutées dans *KRAS*, *BRAF*, *NRAF*, *PIK3CA7* ou *PTEN* ne répondent pas au traitement par le Cetuximab et ont donc un mauvais pronostic.

Même en l'absence de mutation en aval de l'EGFR, il a été montré que des mutations du récepteur lui-même pouvaient expliquer la non-réponse au traitement. Une mutation du gène *KRAS* est présente dans environ 40% des mCRC et le traitement au Cetuximab n'a alors aucun effet. Chez ces patients l'effet du Bevacizumab est également réduit, l'augmentation de la survie totale ou sans progression étant plus faible que dans la population non-mutée. Notre projet cherche donc à rétablir la sensibilité de cellules mutées *KRAS* au traitement par le Cetuximab *in vitro* et d'identifier les cibles intracellulaires impliquées. Pour cela, nous avons réalisé un criblage fonctionnel à l'aide de la banque

HUSCIV dans la lignée HCT116 et en présence du Cetuximab. La prolifération des cellules est détectée à l'aide d'un colorant fluorescent (CMRA). La sélection à l'aide du CMRA comme marqueur fluorescent est bien documentée et permet de suivre chaque division cellulaire avec une bonne résolution. Dans notre cas, un tri par Cytométrie en flux est réalisé et les cellules dont la prolifération est bloquée sont sélectionnées.

A ce jour, nous avons pu réaliser deux tours de sélection. Des tests de cytotoxicité réalisés sur des populations sélectionnées ont montré un enrichissement en scFv ayant un effet cytostatique. En effet, nos résultats montrent une baisse significative (~10%) de la viabilité cellulaire après le 2^{ème} tour de sélection. Ce chiffre n'atteint pas la baisse de viabilité d'environ 50% observée dans la lignée DLD1 sensible au Cetuximab. Toutefois, cette observation indique l'évolution du phénotype qui tend vers une sélection de scFvs d'intérêt et suggère qu'au moins un ou plusieurs tours de sélection supplémentaires seront nécessaires avant de confirmer ces résultats.

Dans ce modèle, notre but est d'identifier les intracops responsables de l'inhibition de la croissance cellulaire ou la mort cellulaire et dont le nombre sera en baisse ou même en disparition lors de la sélection (épuisement). Ainsi, un séquençage profond et à haut débit des populations de scFv évoluant tout au long de l'expérience est nécessaire afin de permettre une analyse statistique. Du fait que les positions variables soient réparties le long de l'ensemble des boucles de CDR, et en raison de la taille des fragments séquencés (~550 paires de base), nous avons opté pour l'identification des scFv par des codes-barres. A ce jour, le séquençage de l'échantillon d'essai a révélé que 43% de nos échantillons contenaient un code barre correct, ce qui correspond à la diversité attendue de 10^6 . Ainsi, 10^8 séquences seront comparées entre les deux groupes (avant et après le traitement par Cetuximab) et des séquences spécifiquement appauvries lors de la sélection seront identifiées.

L'approche introduite ici est différente de toutes les études existantes en ce qu'elle utilise des banques « naïves ». En tant que telle, la stratégie est analogue à des approches génétiques classiques à l'exception qu'elle agit au niveau de la protéine. En effet, l'intracops joue le même rôle que les mutations. Néanmoins, parce que l'intracops ne modifie pas directement sa cible mais module sa fonction, il est également analogue à un médicament. Toutefois, contrairement aux approches typiques utilisant les shRNA et siRNA, l'identité de la cible ne peut être déterminée à partir d'une séquence d'anticorps. En résumé, l'approche décrite ici représente un outil puissant, non seulement pour répondre à la diversité du protéome, mais aussi pour l'étude de messagers secondaires et du métabolisme dans les cellules. En tant que tel, et par rapport à d'autres approches à grande échelle, elle représente une voie simple pour la découverte de molécules thérapeutiques potentielles.

Table of contents

Objectives	17
Introduction	19
1. Intrabody concept	19
1.1 Structure and function of antibodies	19
1.2 Antibody engineering	25
1.3 Intracellular immunization	27
1.4 Selection systems	31
1.5 Recombinant antibody libraries	33
1.6 Applications	37
2. Protein interference	41
2.1 Loss-of-function phenotypic analyses at the mRNA Level	41
2.2 Loss-of-function phenotypic analyses at the protein Level	43
3. Model 1: The regulation of mast cells responses	47
3.1 Mast cells: Origin and functions	47
3.2 Signalization pathways	49
3.3 Mast cell degranulation	51
3.4 The RBL-2H3 cell line	51
4. Model 2: Colorectal cancer and the resistance to Cetuximab	53
4.1 Introduction	53
4.2 The adenoma–carcinoma sequence model	53
4.3 Current therapies for colorectal cancer	55
4.4 The EGFR pathway and its implication in cancer therapy	61
4.5 Treatment directed towards the Epidermal Growth Factor Receptor	61
4.6 The KRAS mutational status	63
4.7 Other pathways activated by EGFR	65
4.8 Treatment directed towards targets downstream of EGFR	65

4.9	Induction of mutations during anti-EGFR treatment	67
Results	69
1.	Phenotypic targeting of mast cells and their role in allergy	69
1.1	Phenotypic screen and target identification	71
1.2	Plasmid library sequences	72
1.3	Retroviral library sequences.....	73
1.4	Characterization of inhibitory intrabody families	73
1.5	Target identification	75
2.	Optimization and expression of the intrabody library	79
2.1	Objectives.....	79
2.2	Design of HUSCIV	81
2.3	HUSCIV expression in eukaryotic cells.....	87
3.	Phenotypic targeting of KRAS mutated colorectal cancer cells	91
3.1	Objectives.....	91
3.2	Experimental set-up	91
3.3	Phenotypic screen	95
3.4	Cytotoxicity experiment	97
3.5	Sequence analysis	99
Conclusions	103
Annex 1: Materials and methods	113
1.	Synthesis and optimization of the scFv library.....	113
1.1	Preparation of chemo-competent E. coli	113
1.2	Preparation of uracil-containing single-stranded DNA (ssDNA)	113
1.3	Annealing the library mutagenic primers to the ssDNA template	113
1.4	Synthesis of covalently closed circular scFv library: extension/ligation	114
1.5	Transformation of bacteria with scFv library by electroporation	114
1.6	Purification of scFv library (MAXI prep)	114

2.	Cell cultures.....	114
2.1	Rat mast cell line RBL-2H3.....	114
2.2	293T (or HEK-T)	114
2.3	Colorectal cancer cell lines: HCT116, CaCo2 and DLD-1	115
3.	Production and purification of antibody fragments	115
3.1	Production of antibody fragments from cytoplasmic extracts of bacteria	115
3.2	Production of antibody fragments from bacterial periplasmic extracts.....	115
3.3	Purification on Nickel resin	115
3.4	Purification on magnetic beads of nickel	115
3.5	Production and immuno-precipitation of cell extracts.....	116
4.	HUSCIV expression in eukaryotic cells.....	116
4.1	Chemicals and enzymes	116
4.2	Plasmids.....	116
4.3	Retrovirus production	116
4.4	Retroviral transfection	117
4.5	Flow cytometry analysis and cell sorting	117
4.6	Cell extracts, scFv precipitation and western blot analysis.....	117
4.7	Fluorescence microscopy	118
5.	Phenotypic targeting of mast cells and their role in allergy	118
5.1	Retrieving scFv sequences.....	118
5.2	Recloning in retroviral vector pMSCV-SN-EGFP	118
5.3	Retrovirus production and transfection.....	118
5.4	Analysis of stable transfectants	119
6.	Phenotypic targeting of <i>KRAS</i> mutated colorectal cancer cells	120
6.1	Retrovirus production	120
6.2	Viral titers.....	120
6.3	Retro-transfection	120

6.4	FACS analysis of HCT116 stained with CMRA.....	120
6.5	Recloning of CMRA+ population	121
6.6	Preparation of barcoded DNA fragments for Illumina sequencing.....	121
Annex 2:	Recloned intrabody family sequences.....	123
Annex 3:	HUSCIV sequences.....	124
References	127
Article Mazuc <i>et al.</i>, 2014	141

Objectives

The main objective of my thesis was to perform a phenotypic screening approach in a colorectal cancer model. Our aim was to restore the sensitivity of KRAS mutated cells to Cetuximab treatment *in vitro* by the identification and inhibition of intracellular targets that may interact with the EGF pathway.

For this purpose, a highly diverse retroviral intrabody library was designed with improved diversity, affinity and stability of scFv fragments inside the cells. This optimized human intrabody library (HUSCI), was used to transfect a KRAS mutated adenoma-carcinoma cell line, and intrabodies of interest were identified by two methods:

- 1) FACS-based screening of cells blocked in their proliferation by means of a fluorescent dye;
- 2) Statistical analysis of declining sequence frequencies, obtained by next generation sequencing of the subsequent intrabody pools;
- 3) The rescue of full intrabody sequences from the highly diverse library, needed for the identification of their targets.

1. Intrabody concept

1.1 Structure and function of antibodies

Upon an infection of a higher organism, two major systems try to counteract the pathogen invasion in the body. Initially, the innate immune system provides an immediate, non-specific response after the invasion of an unknown pathogen (Palm and Medzhitov, 2007). If the infection continues, the response of the adaptive immune system is initiated. The adaptive immune response is antigen-dependent and specific and is developed during the lifetime of an individual. It is based on a highly advanced interplay between a variety of different cell types and molecules among which lymphocytes and antibodies are the key elements. There are two types of lymphocytes, T cells and B cells. T cells are responsible for cellular immune response, while B cells are essential for the humoral immune response. In both cases the immune response is stimulated by specific antigens which can range from small chemical structures, through sugars, lipids, peptides and nucleotides to large molecules, such as proteins, protein complexes or even whole organisms like bacteria, viruses and parasites (Delves and Roitt, 2000).

During adaptive immune response, B cells are providing both specific and long-lasting protection against a diverse range of pathogens (Harwood and Batista, 2010). Once an antigen is bound to the surface of a B cell via the B cell receptor (BCR), it gets internalized and processed into peptides (Parker, 1993). The processed antigen is returned to the B cell surface and presented via the Major Histocompatibility Complex (MHC) II for antigen specific helper T cell (T_H cell) recognition (Noelle and Snow, 1992). The physical contact between B and T_H cell triggers the production of specific interleukins including IL-4, IL-5 and IL-6 which in turn activate B cells (Parker, 1993). IL-4 activates clonal proliferation, while IL-5 and IL-6 induce the differentiation of the B cell. The activated B cells either differentiate into memory B cells providing long-term protection against secondary infection or into antibody secreting plasma cells (Reth, 1992).

In the immune system antibodies are responsible for recognition and binding of foreign structures. They protect the host in different ways. First, they neutralize toxic effects of pathogens by binding pathogens and hence preventing infection. Second, they facilitate uptake of the pathogen by phagocytosis and third, they recruit natural killer (NK) cells by binding to the surface of a pathogen

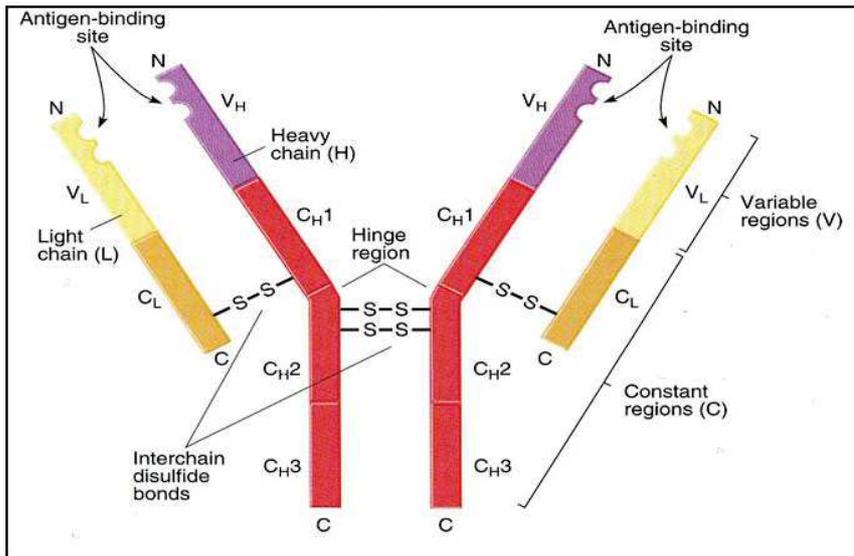


Figure 1: General structure of an immunoglobulin in higher vertebrates illustrated in this case by IgG. (Thompson et. al, Austin Peay State University).

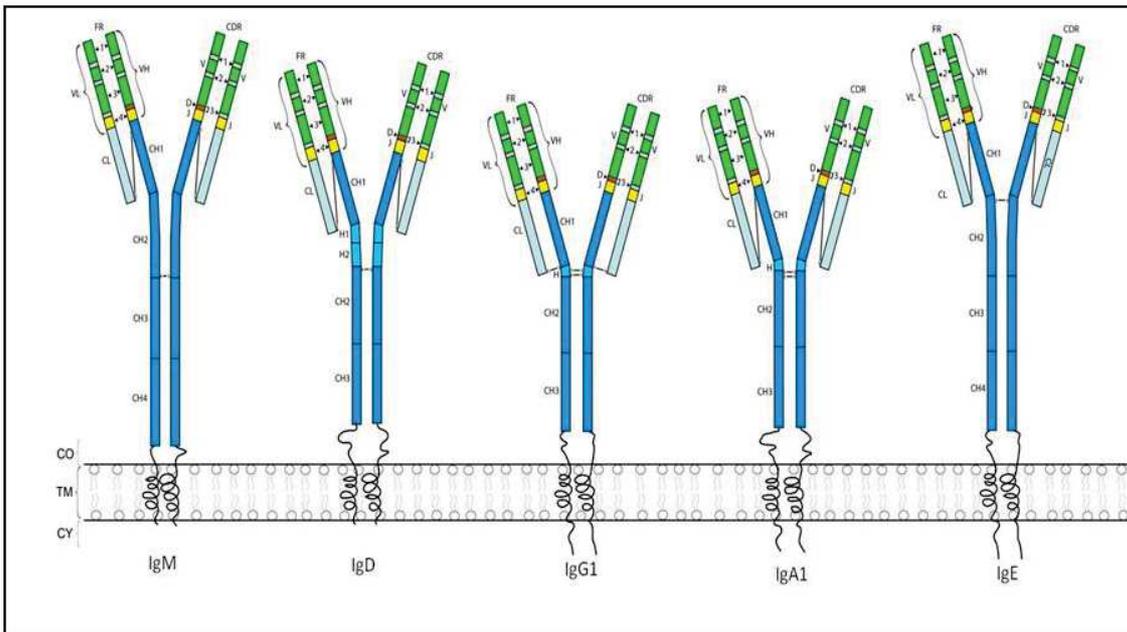


Figure 2: Membrane-anchored forms of the immunoglobulin isotypes are encoded by a cluster of immunoglobulin heavy chain constant region genes (IMGT Education).

enabling NK cells to destroy the antibody-coated cells in a process called antibody-dependent cell-mediated cytotoxicity (ADCC).

Antibodies are capable of recognizing an enormous variety of different antigens. However, they do not recognize whole antigens, but parts of it referred to as epitopes. The diversity of antibodies is based on their modular architecture making it possible to generate up to 10^{11} or more different antibodies (Willis *et al.*, 2013). In humans and closely related species the basic antibody molecule (~150 kDa), also referred to as immunoglobulin (Ig), consists of two identical heavy (H) chains and two identical light (L) chains. Two types of light chains (κ and λ) and five types of heavy chains (μ , δ , γ , ϵ and α) can be distinguished. Heavy and light chains are linked covalently by disulfide bonds (Porter, 1963) (Figure 1). Both, the heavy (~50 kDa) and light chain (~25 kDa) are composed of two regions with distinct variability of their amino acid sequence (Hiltschmann and Craig, 1965): the variable (V) region at the N-terminus and the constant (C) region at the C-terminus (Cohen and Milstein, 1967). Each light chain consists of one variable domain (VL) and one constant domain (CL). Each heavy chain, however, consists of one variable domain (VH) and up to four constant domains (CH). The so-called hinge region is located between CH1 and CH2 constant regions of the heavy chain forming the typical Y-shaped structure of an antibody. Antibody molecules can be cleaved in distinct fragments after incubation with the protease papain. It cuts the molecule in the hinge region to yield the Fab (fragment antigen binding; ~45 kDa) and the Fc (crystallizable fragment; ~50 kDa) part of the molecule. The Fab part retains the antigen binding activity while the Fc part contains most of the constant region of the heavy chains (Delves and Roitt, 2000).

Five classes of immunoglobulins can be recognized; IgM, IgD, IgG, IgE and IgA. The different heavy chains that define these classes are known as isotypes and are designated by the lower-case Greek letters μ , δ , γ , ϵ , and α . At the membrane of a mature B cell, IgMs are the first type of immunoglobulins which are expressed, followed by IgDs (Harriman *et al.*, 1993). The expression of these two classes is mainly due to a differential splicing of a primary mRNA. IgMs comprise the heavy chain constant region $C\mu$ while IgDs contain $C\delta$ (Figure 2).

During a somatic recombination process called class switch recombination (CSR) (Li *et al.*, 2004), the gene coding for CH is changed from $C\mu$ to $C\gamma$, $C\epsilon$ or $C\alpha$, allowing for switch of isotype from IgM to either IgG, IgE or IgA, respectively. The five isotypes also show structural differences including the number and location of disulfide bonds, the number of C domains and the length of the hinge region (Figure 2).

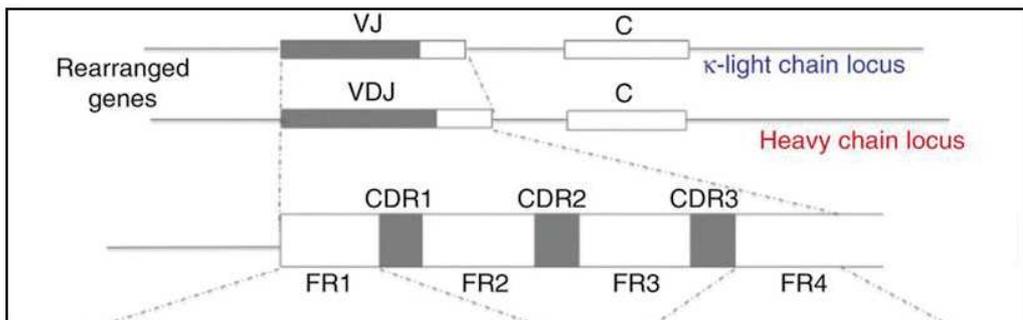


Figure 3: Schematic illustration of VH and VL domains (Muzard et al., 2013).

This hinge region connects the CH1 and CH2 domain and is responsible for the flexibility of an antibody molecule (Padlan, 1994). Since the length of the hinge region varies in the different isotypes, the distance between the antigen binding sites differs. While IgD, IgG and IgA class antibodies have a hinge region, IgM and IgE class antibodies are lacking one but contain an extra heavy-chain domain. These differences in the heavy chain composition are responsible for distinct characteristics and effector functions of each isotype. As described above, IgMs are the first antibodies to be produced and hence are responsible for activation of phagocytosis. IgA class antibodies are mainly found in mucosal surfaces. IgEs are involved in defense against parasites and play a major role in common allergic diseases, whereas IgD isotypes exist only in minor amounts as antibody in serum and its function is still unsolved (Reth, 1992). During immune response of higher vertebrates IgGs are the most abundant antibodies in serum (Cohen and Milstein, 1967).

In humans, IgGs can be further divided into four subclasses: IgG1, IgG2, IgG3 and IgG4. The IgG subclasses are named in order of abundance in serum: IgG1 70 %, IgG2 20 %, IgG3 7 % and IgG4 3 % (Ochs and Wedgwood, 1987). The structural differences among IgG subclasses lie in the length of the hinge region and in the distribution of the disulfide bonds (Ochs and Wedgwood, 1987). Antibodies specifically recognize and bind antigens through the variable region located at the N-terminus of the molecule. The variable region differs extensively between antibody molecules, but the sequence variability, however, is not distributed equally throughout the variable regions but concentrated in three so called hypervariable regions. When VH and VL domains are paired in the antibody molecule, three hypervariable loops from each domain are brought together creating a single hypervariable site: the antigen-binding site (Figure 1). The hypervariable loops form the complementarity determining regions (CDR1, CDR2 and CDR3), and are encoded by a genetic recombination of VJ in the case of a light chain region and VDJ in the case of heavy chain regions. The CDRs are flanked by less variable and highly conserved regions which are called framework regions (FR) (Kabat and Wu, 1991) (Figure 3). However, mutations in FRs are also known to influence binding since they play an important role in structure preservation, folding yield, stability and even direct interaction of the antibody with the antigen (Padlan, 1994) (David *et al.*, 2007).

The CDRs of the light chain are roughly 6 to 10 amino acid residues in length, those of the heavy chain are roughly 5 to 15 amino acid residues in length (Kabat and Wu, 1991) (Johnson and Wu, 2000), but they show wide variations in length between species (Padlan *et al.*, 1995) (Collis *et al.*, 2003), especially in the CDR3 of heavy chains (Padlan and Kabat, 1991) which often plays a central role in antigen binding. Since antigen-binding is mediated through CDRs from both VH and VL domains,

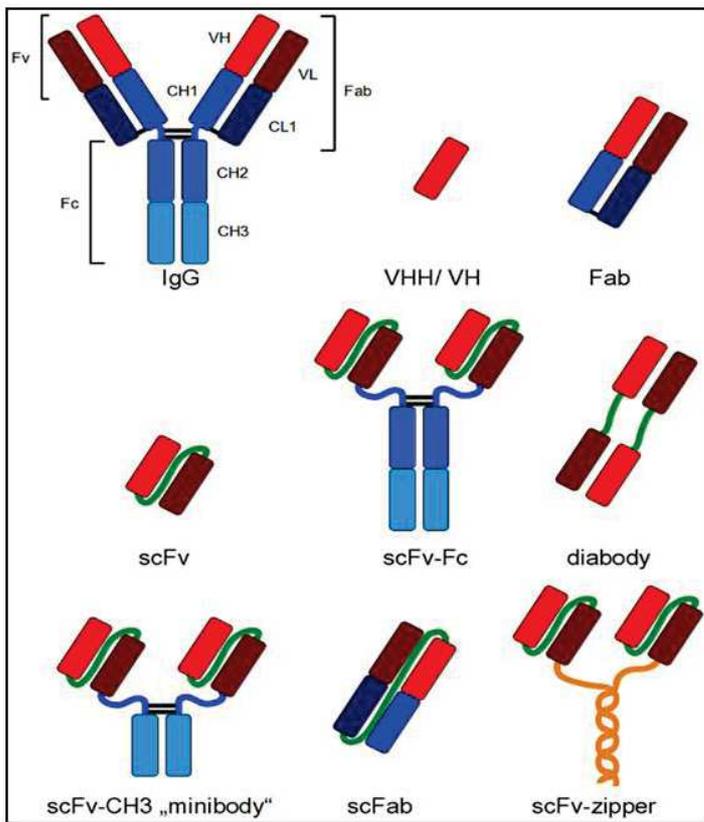


Figure 4: Schematic illustration of different antibody formats derived from conventional antibodies (Frenzel et al., 2013).

it is the combination of heavy and light chain which determines the antigen specificity. As a result, antibodies naturally are bivalent with two antigen-binding arms of identical specificity (Delves and Roitt, 2000). Antibody-antigen interactions are non-covalent bonds such as hydrogen bonds, salt bridges and van der Waals bonds (Absolom and van Oss, 1986) and the strength of binding is called affinity. The presence of two identical antigen-binding sites allows antibodies to bind simultaneously to two identical antigens and hence increases the total strength of the interaction.

1.2 Antibody engineering

Recombinant antibodies are antibodies that are not produced from their natural locus in B-cells. Nowadays, recombinant antibodies are created by genetic engineering leading towards recombinant antibody libraries based on diversified antibody gene segments. It is possible to create recombinant antibodies which usually are difficult to obtain by the immune system, e.g. due to transient conformational change of an antigen after cofactor binding (Nizak *et al.*, 2003) (Figure 4). The smallest antigen binding fragment of immunoglobulins maintaining its complete antigen binding site is the Fv fragment, which consists of the variable regions (VH + VL). Naturally, VH and VL domains are non-covalently associated via a hydrophobic interaction and tend to dissociate (Winter and Milstein, 1991). Often, a soluble and flexible amino acid peptide linker ((Gly4Ser)₃) is used to connect the V regions to a scFv (single chain fragment variable, ~25 kDa) fragment for stabilization of the molecule, hence representing the minimal recombinant antigen-binding fragment of antibodies (Hudson, 1998). It consists of the two variable domains, VH and VL. As scFvs are relatively small and unglycosylated, they can be produced in heterologous expression systems like bacteria or lower eukaryotes (Skerra and Plückthun, 1988). Moreover, their small size has been suggested to permit binding to cryptic epitopes not accessible to full-sized mAbs (Ward *et al.*, 1989).

But due to fast absorbance and excretion by the kidney, scFvs demonstrate short circulating half-lives in the organism (Sanz *et al.*, 2005). Furthermore, therapeutic applications or diagnostic applications are still hampered by stability problems since scFvs still have a strong tendency to aggregate due to hydrophobic interactions in a hydrophilic environment (Nieba *et al.*, 1997).

In some species, like the family of Camelidae, light chains are absent in some of the Ig and thus their combining sites consist of only one domain, termed VHH (Hamers-Casterman *et al.*, 1993). Because they are relative small, they outperform other antibodies in terms of stability, resistance to aggregation, refolding properties, expression yield, DNA manipulation, library construction,

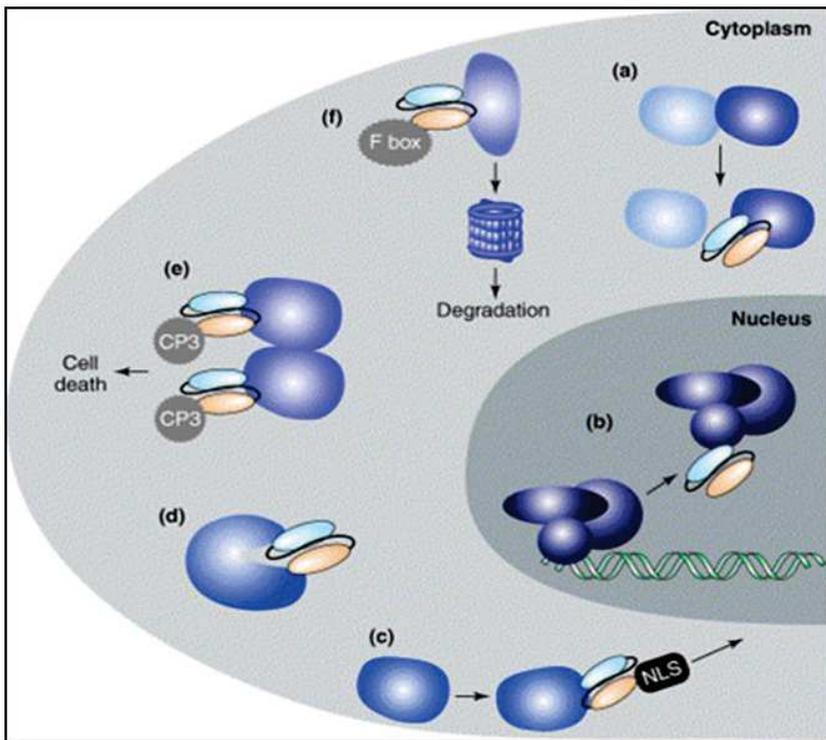


Figure 5: Intrabodies modes of action (Lobato et al., 2003)

and three-dimensional structure determination. Many of these are desired in applications involving antibodies. However, the non-human nature of VHHs limits their use in human immunotherapy because of the immunogenicity issue. In this respect, human VHs are ideal candidates for therapeutic applications because they are expected to be least immunogenic. However, due to the missing VL interaction they have an exposed hydrophobic surface and tend to aggregate (McGregor *et al.*, 1994). Thus, attempts were made previously to obtain monomers, i.e. human VHs suitable for antibody applications. Such VHs also displayed other useful properties typical of VHHs, such as high expression yield, and high renaturation yield. Synthetic libraries built on these VHs as library scaffolds should serve as a promising source of therapeutic proteins.

Besides scFvs and VHs, Fab (~50 kDa) fragments are also commonly used recombinant antibody fragments. Fabs can be recombinant or generated by proteolytic digestion of full length antibodies and comprise the VH and CH1 domains of the immunoglobulin heavy chain and the VL and CL domains of the light chain. Fab fragments are used for diagnostic imaging (Behr *et al.*, 1995) and *in vivo* imaging (Elsässer-Beile *et al.*, 2009). Recently, 3 Fabs have obtained FDA approval as for example Ranibizumab, an anti-VEGF antibody used in the treatment of certain forms of age-related macular degeneration (Stewart, 2014).

1.3 Intracellular immunization

Intracellular antibodies or 'intrabodies' are antibody fragments that are expressed within the cell and directed against intracellular proteins. In this way they can interfere and inhibit cellular processes inside the cell in a number of ways (Figure 5), affording them great potential for the use in target validation. Intrabodies can inhibit an enzymatic activity directly, or interfere with protein-protein interactions, thus disrupting signaling pathways. They can also be used to displace a protein from its site of action. The fusion of intracellular localization signals, such as a nuclear localization signal (NLS) or a retention signal for the endoplasmatic reticulum (ER) can be used to re-direct the intrabody and its target antigen to specific locations within the cell. For instance, an scFv directed against the ErbB-2 receptor and designed to prevent transit through the ER was shown to down-regulate the surface expression of ErbB-2 and consequently, to considerably affect growth factor signaling (Graus-Porta *et al.*, 1995).

In another elucidating study on the action of intrabodies, the authors compared the inhibition of Ras function with a number of different scFv fragments. Their ability to inhibit Ras function was then characterized *in vivo* using a cell proliferation assay and *in vitro* measuring GTPase activity.

Although none of the selected scFvs were able to inhibit Ras function *in vitro*, it emerged that a number of them still functionally inhibited Ras *in vivo* by diverting it from the plasma membrane and sequestering it the nucleus or by the formation of aggregates. These studies illustrate the potential of intrabodies to validate targets in cellular assays (Lener et al., 2000).

The work of Lecerf et al (Lecerf *et al.*, 2001) describes the use of human scFvs directed against huntingtin protein to interfere with the formation of intracellular aggregates characteristic of Huntington's disease. The inhibition of protein aggregation required the fusion of a nuclear localization signal (NLS) to the scFv, which was subsequently able to retarget the huntingtin derivative to the nucleus. The authors suggest that the binding event, by maintaining huntingtin protein in a soluble state, favors normal cellular protein turnover rather than aggregation. This 'solubilizing' property of scFvs could be interesting for a range of neuro-degenerative diseases caused by abnormal protein aggregation.

Also in the field of "HIV therapy" scFv were applied. The scFv 105 against the Env protein (gp120) was shown to inhibit the proteolytic processing of the precursor protein gp160 in the ER and decrease the infectivity of HIV virions released by the cells (Marasco *et al.*, 1993). Various scFvs have been selected to target Tat, another protein essential for the life cycle of HIV.

Tat is required for the transactivation of the HIV-LTR. Once the scFv was bound to Tat, it blocked the latter's nuclear function. Interestingly, nuclear targeting of the scFv was not required for this effect (Mhashilkar *et al.*, 1995). This suggests that the scFv-mediated effects were exerted by sequestering Tat in the cytoplasm, rather than interfering mechanistically with its nuclear function. Similarly, scFvs directed against HIV's Rev were shown to prevent the cytoplasmic nuclear shuttling of Rev, a regulatory RNA binding protein. Curiously, the *in vitro* affinities of the scFvs did not always correlate with the observed performance in the cellular system; some scFvs with lower affinity displayed a more potent inhibitory effect *in vivo*. This demonstrates the importance of considering the different requirements of the intracellular environment for the functionality of an intrabody.

As a final example, our group has shown the efficient inhibition of the Spleen tyrosine kinase (Syk) with intrabodies directed against its SH2 domains. The cytoplasmic expression of anti-Syk intrabodies in a mast cell line affected protein-protein interactions of Syk with its cellular targets and inhibited mast cell activation (Dauvillier *et al.*, 2002) (Mazuc *et al.*, 2008). Subsequently, these anti-Syk intrabodies have been used in a high throughput screening assay to identify non-enzymatic inhibitors of Syk and allergy (Villoutreix *et al.*, 2011).

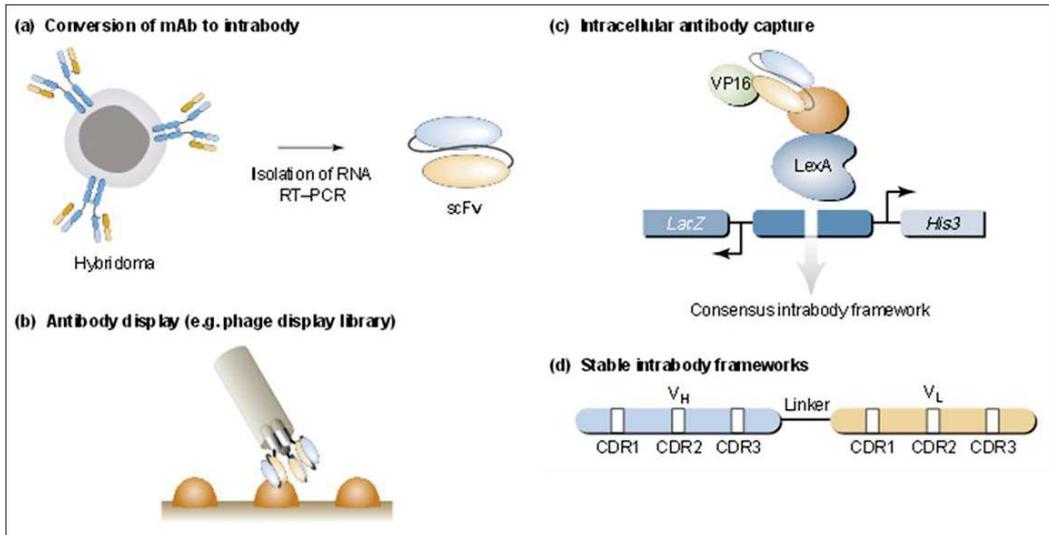


Figure 6: Methods for making and selecting intrabodies (Lobato et al., 2003).

Despite these encouraging results, the use of antibodies inside the cell is troubled with difficulties regarding stability and solubility. This has limited their widespread use to date. Disulfide bridges contribute greatly to the stability of the intrabody. In addition to the inter-molecular disulfide bridges important for the stability of the full size intrabody, intra-molecular disulfide bridges are crucial for the structural integrity of the intrabody (Biocca *et al.*, 1990) (Cattaneo and Biocca, 1999). Therefore, only a small proportion of intrabodies are functional under the reducing intracellular conditions, while the majority of intrabody frameworks become unstable, insoluble and therefore non-functional. However, as the potential of intrabodies for applications inside the cell is becoming increasingly recognized, intrabody engineering and selection procedures have been developed to isolate functional intrabodies. Further development of these technologies in large-scale libraries will greatly facilitate the exploitation of intrabodies.

1.4 Selection systems

A straightforward approach to intrabody isolation is derivation from the V regions of a high-affinity monoclonal antibody (mAb) (Figure 6a). The VH and VL can be amplified by reverse transcriptase (RT) – PCR of mRNA from the original hybridoma and then be cloned as a scFv (Marasco *et al.*, 1993). Alternatively, one of the *in vitro* display systems, such as phage (Figure 6b), bacterial or yeast surface or ribosome display techniques, can be employed to screen scFv libraries with the desired antigen to select specific scFv that can subsequently be tested as intrabodies (Wörn and Plückthun, 2001).

The problem with these approaches is that no allowance is made for the inability of some scFv to function in the cellular milieu. As a result, the scFv can be nonfunctional and show poor expression levels, low solubility and/or a short half-life. Although screening high-diversity scFv libraries should yield low numbers of scFv that will work in cells, these intrabodies could be quickly identified using ‘intracellular antibody capture’ (IAC, Figure 6c) (Tse *et al.*, 2002), which combines a first round of *in vitro* screening of diverse phage scFv libraries with a second intrabody – antigen interaction screen in yeast (Visintin *et al.*, 1999) to determine the best *in vivo* interactions. It is also possible to directly isolate single domains that bind to target proteins within the reducing cellular environment (Tanaka and Rabbitts, 2012). In addition, this method also allows a consensus scaffold to be defined, with improved solubility and expression properties *in vivo* for the development of improved intrabody libraries (Tanaka and Rabbitts, 2003).

Several other techniques have been developed, which rely on a pre-determined ability of intrabody fragments to fold adequately and to remain stable in cells using optimized scFv frameworks (Figure 6d). These include fragments that tolerate the absence of intra-chain disulfide bonds (Wörn and Plückthun, 1998), fragments made by CDR grafting onto a stable intracellular scFv (Ohage *et al.*, 1999) (Philibert *et al.*, 2007) and fragments produced by rational stability engineering of variable domains (Desiderio *et al.*, 2001). These approaches promise to deliver effective intrabody libraries. Future adaptations of all these methods will allow the development of rapid-throughput approaches to the production of panels of intrabodies against protein targets in disease-state cells and in functional genomics applications.

1.5 Recombinant antibody libraries

As described above, recombinant antibodies are frequently derived from antibody libraries. The most prominent Fab library is the HuCal library (Knappik *et al.*, 2000) which is a synthetically generated library with several randomized positions in the 6 CDR-regions. To select desired antibodies from such libraries, display technologies such as bacterial display, ribosome display and phage display have proven to be powerful tools. With these methods, recombinant binding molecules can be selected *in vitro* from comprehensive libraries derived synthetically (Knappik *et al.*, 2000) or from immunized (He *et al.*, 1999) or naïve animals (de Haard *et al.*, 1999).

Naïve antibody libraries

A truly 'naïve' antibody library would be made from IgM bearing cells prior to exposure to antigen. In common usage it may simply mean that the individuals whose intrabody-expressing cells were used to generate the libraries were not specifically immunized against particular antigens. In practice, lymphocytes are isolated from a number of individuals and heavy and light chains are amplified for insertion into expression vectors for display.

Immune antibody libraries

Immune antibody libraries are created starting with lymphocyte RNA from individuals previously exposed to the desired antigen. Such libraries are biased toward high affinity binders for the desired antigens since they contain heavy and light chains that have undergone *in vivo* affinity maturation. Such libraries can be of a smaller size than 'universal' libraries. Of course, poorly immunogenic antigens, including very highly conserved antigens, are problematic and the purposeful exposure to antigens is limited to non-human subjects.

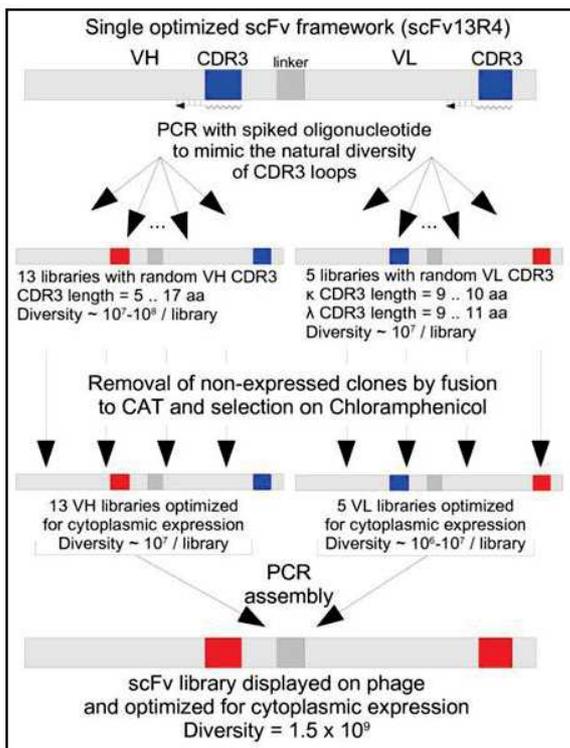


Figure 7: Development of a scFv library optimized for intracellular expression (Philibert et al., 2007)

Semi-synthetic antibody libraries

Semi-synthetic libraries have combinations of natural and synthetic antibody diversity; they are often created to increase natural diversity while maintaining a certain level of functional diversity. For example, libraries created by shuffling natural CDR regions or by introducing naturally rearranged and highly functional CDR3 sequences from human B-cells with synthetic CDR1 and CDR2 diversity.

De Kruif *et al.* used 49 germline VH genes in combination with synthetic CDR regions and seven light chains to create a human scFv phage display library from which phages were isolated using 13 different antigens, including ones that could distinguish between two highly related homeobox proteins (de Kruif *et al.*, 1995). Another research group paired a single VH chain wherein random mutations were introduced using partially degenerate primers and PCR with a single VL chain to create a human phage display library. Hoet *et al.* combined synthetic diversity in CDR1 and CDR2 with naturally occurring sequences to create human Fab libraries from which high affinity antibodies were obtained. By using the heavy chain CDR3 regions from 35 donors they were able to capture both the sequence and length diversity naturally present in CDR3. Specific amino acids in CDR1 and CDR2 were varied in such a way as to increase the likelihood of creating higher affinity Fabs from the phage display library (Hoet *et al.*, 2005).

Synthetic antibody libraries

Synthetic antibody libraries are constructed entirely *in vitro* using oligonucleotides that introduce areas of complete or tailored degeneracy into the CDRs of one or more V genes. Synthetic diversity bypasses the natural biases and redundancies of antibody repertoires created *in vivo* and allows control over the genetic makeup of V genes and the introduction of diversity. Knappik *et al.* used consensus frameworks and a cassette approach for introducing diversity into the CDR3 regions of both heavy and light chains. Seven master heavy chain genes and seven master light chain genes were synthesized based on consensus germ line sequences. Codons were optimized for expression in *Escherichia coli* and certain problematic residues were eliminated altogether. All 49 combinations of heavy and light chains were cloned into a phagemid vector. Both the VH and VL CDR3 regions were replaced by a mixture of trinucleotides skewed toward amino acids naturally occurring in human CDR3 sequences. In this way, diversity can be both created and controlled, depending on the mixture of trinucleotides in the mutagenesis cassette (Knappik *et al.*, 2000). Within our team, a synthetic library of scFvs optimized for intracellular expression was constructed by using the consensus sequence 13R4 as a framework (Figure 7). The diversity was introduced at the level of the CDR3 regions (Philibert *et al.*, 2007) and the template (scFv13R4) was optimized by molecular

evolution during previous works, which aimed for optimal expression and activity in the cytoplasm (Martineau *et al.*, 1998a).

For that purpose, a scFv fragment that binds and activates an inactive mutant β -galactosidase was isolated. The gene encoding the scFv fragment was subjected to random mutation *in vitro* by error-prone polymerase chain reaction, and co-expressed with the mutant β -galactosidase in lac^- bacteria. By plating on limiting lactose, they selected for intrabody mutants with improved expression, and after four successive rounds of mutation and selection, an intrabody fragment was isolated that is highly expressed in the bacterial cytoplasm. Analysis of the mutant intrabody fragments revealed that the disulphide bonds are reduced in the cytoplasm and that the fragments could be denatured and renatured efficiently under reducing conditions *in vitro*.

This shows that with a suitable method of screening or selection, it is possible to make folded and functional intrabody fragments in excellent yield in the cytoplasm. These results were also confirmed in yeast (Visintin *et al.*, 1999), in plants (Phillips *et al.*, 1997) and in mammalian cells (Lener *et al.*, 2000) (Sibler *et al.*, 2003).

1.6 Applications

The importance of the various intrabody modalities shown in figure 5 is that specific properties of target proteins can be exploited to provoke a specified response and to investigate or target a particular protein domain. There are several studies describing the potential applications of intrabodies in AIDS, including a scFv derived from a human mAb that binds glycoprotein 120 (gp120), the envelope protein of HIV-1 (Marasco *et al.*, 1993). This intrabody not only binds to gp120 but also retains it in the ER of infected cells, thereby preventing the formation of enveloped virus. A similar strategy has also been used to prevent surface expression of the chemokine receptor CCR5, thereby protecting T cells against viral infection (Steinberger *et al.*, 2000). Other attempts have been made to inhibit the infectivity of HIV-1, by generating scFv that target viral proteins mainly implicated in replication. In the field of cancer, intrabodies have been used to modulate the expression of proteins upregulated in tumors, such as ERBB2 in breast and ovary tumors (Alvarez *et al.*, 2000), interleukin-2 receptor α in some leukemia (Richardson *et al.*, 1995), cyclin E in breast cancer (Strube and Chen, 2002), and epidermal growth factor receptor (EGFR) in glioblastoma and epithelial cancers (Hyland *et al.*, 2003) (Jean *et al.*, 1996). In all these cases, appropriate cellular localization signals were attached to the intrabodies to reduce the activity of tumor-related proteins by altering their location (Fig. 5c).

Oncogenic and tumor suppressor proteins, such as p53 (Cohen *et al.*, 1998) and RAS (Tanaka and Rabbitts, 2003), which are mutated in a large number of tumors, and fusion proteins resulting from chromosomal translocations, such as BCR –ABL (Tse *et al.*, 2002) in leukaemia, are good candidates for intrabody therapy. Intrabody-mediated therapies might also be a viable option for neurodegenerative diseases (e.g. Alzheimer's, Parkinson's or prion diseases), as well as for diseases caused by mutated proteins or infectious agents. For instance, scFv have been designed for use in Huntington's chorea (Lecerf *et al.*, 2001), in skin transplants (Mhashilkar *et al.*, 2002) and against the Tau protein in Alzheimer's disease (Visintin *et al.*, 2002).

In addition to targets in diseased cells, intracellular antibodies could play an important role in target validation in functional genomics. Intrabodies can be used to study the function of the thousands of novel products identified from the Human Genome Project and other sequencing programs. Strategies that use intrabodies to ablate the function of a protein in tissue culture and during development (phenotypic knockout), will allow determination of the physiological and pathological relevance of this protein. As a result, proteins with therapeutic potential can be selected.

Furthermore, the identified intrabodies could be the basis for therapeutic compounds. The tyrosine kinase Syk, for example, has shown a high potential for the discovery of new treatments for inflammatory and autoimmune disorders. Pharmacological inhibitors of Syk catalytic site have been developed but showed limited specificity towards Syk (Brasemann *et al.*, 2006). Our lab opted for the design of drug-like compounds that could interfere with Syk while maintaining its kinase activity. For this challenging task, the lab used the potential of intracellular antibodies for the modulation of cellular functions *in vivo*, combined to structure-based *in silico* screening. Then, an intrabody displacement assay was used in order to screen for functional mimics and found 10 compounds that inhibited degranulation with IC₅₀ values ≤ 10 μM. These compounds could be a lead towards the development of new classes of non-enzymatic inhibitors of Syk and for drug discovery in the field of inflammation (Mazuc *et al.*, 2008) (Villoutreix *et al.*, 2011).

2. Protein interference

During the past decade, the complete genomes of more than 140 different organisms have been sequenced and made available in databases (Benson *et al.*, 2003). These databases provide extremely useful collections of organized, validated data, which are indispensable for genomics and proteomics research and the drug discovery process. The human genome encodes approximately 25,000 proteins. Considering that about 60% of genes are alternatively processed for an average of 2-4 messages per gene, and that greater than 200 post-translational modifications are known, then the protein diversity available to organize a complex organism is vast. Yet we know the function of only a minority of those proteins.

Differential analyses of pathogenic and healthy states of organisms or isolated cells provide a picture of genes and gene products that are related to, or actually responsible for, defined diseases. The challenge today is to understand in detail the function and interplay of the numerous proteins in different organisms, tissues, cell types and conglomerate protein complexes. Among the most effective ways to study the function of a given protein in the context of the living cell or organism is the application of a small-molecule drug that exhibits high specificity, affinity and inhibitory activity for the protein under investigation. However, because such inhibitors are available only for a minority of the estimated total number of proteins of higher vertebrate organisms (Harrison *et al.*, 2002), (Claverie, 2001), protein function is most commonly studied by loss-of-function phenotypic analysis.

2.1 Loss-of-function phenotypic analyses at the mRNA Level

Most traditional approaches for this purpose usually rely on observation of phenotypic alterations of a cell or organism as a consequence of alteration of its genetic information. In general, this is achieved either by transgenic knockout technologies (Shashikant and Ruddle, 2003) or by dominant negative expression of a protein or a mutant derivative. However, genome manipulation is a time-consuming and expensive approach, requiring invasive intervention.

A less laborious alternative is to gain functional information by targeted mRNA destruction of the gene of interest. This can be achieved, for example, by antisense oligodeoxynucleotides (ODNs) (Uhlmann *et al.*, 2000) or chemically modified oligonucleotides (Kurreck, 2003) with nucleotide sequences that are complementary to the mRNA to allow sequence-specific hybridization. Protein production is then blocked either by inhibition of ribosome scanning of the mRNA or by activation of

endogenous RNase H, which recognizes these heteroduplexes and hydrolyses the mRNA part. Problems associated with the antisense approach are that many ODNs often exhibit an intolerable degree of toxicity, their target sequences on mRNAs may be inaccessible due to bound proteins, or because the mRNA is engaged in higher-order structures (Branch, 1998).

Other options employed for similar purposes are intracellular ribozymes (Castanotto *et al.*, 2002) or DNAzymes (Cairns *et al.*, 2002). Unlike ODNs, ribozymes have the advantage of cleaving the target mRNA with multiple turnover, while their mechanism of recognition of their target mRNA sequence also operates through simple Watson–Crick pairing. As enzymes that cleave phosphodiester bonds they are independent of the host-cell's endogenous RNase activity. Several examples have shown that intracellularly expressed ribozymes can efficiently down-regulate the expression of proteins; they have been extensively reviewed elsewhere (Rossi, 1995). With regard to cleavage-site selection, ribozymes and DNAzymes face similar problems to ODNs, and some impressive attempts have been successfully undertaken to overcome these obstacles (Chen *et al.*, 1997).

In the past few years, another extremely versatile method for silencing genes on the mRNA level has become available, in the form of short interfering RNAs (siRNAs). siRNAs are RNA double strands of 21–22 nucleotides in length that can downregulate the expression of eukaryotic genes with complementary sequences by utilizing the RNA-induced silencing complex (RISC) protein components of the RNA interference (RNAi) machinery (Tijsterman and Plasterk, 2004). Short interfering RNAs have emerged as a powerful laboratory tool for knocking down gene expression in various cells and organisms, because their design is simple and because they can be easily obtained by standard RNA synthesis, thus allowing straightforward analysis of biological functions of specific genes. Their application potential is wide and, like intracellular ribozymes, siRNAs can be endogenously expressed in a variety of cells.

2.2 Loss-of-function phenotypic analyses at the protein Level

While all these approaches have proved priceless as tools for functional genomics, they share certain disadvantages associated with the alteration of the amount of an expressed protein in the context of its natural functional network in a cell, tissue or organism. Alteration of the genetic information of an organism often has secondary effects on the expression pattern of other genes in a somewhat unpredictable fashion. Also, siRNAs sometimes only give partial knock-down of their target protein or can result in the undesired induction of interferon response (Kim *et al.*, 2004), which may hamper an unbiased analysis of gene function. Specific modulation of gene function at the protein level is

therefore still highly desirable. The post-genomics era and the need to develop novel pharmaceuticals have created a growing demand for specific ligands and inhibitors that will act directly on the protein or a defined protein subdomain without altering the genetic or mRNA status of an organism (Stockwell, 2000). Direct inhibition of a protein allows immediate insight into questions such as drugability, or the functional role of sub-domains or post-translational modifications. The analysis of gene function on the protein level requires direct recognition and inhibition of protein targets by inhibitory molecules that need to fulfill certain criteria: they should be routinely obtainable and applicable independent of the target, and act at low concentrations, with high specificity and in an intracellular context.

A class of molecules that fulfills these requirements is nucleic acid ligands, or aptamers. Aptamers are short, single-stranded oligonucleotides that fold into distinct three-dimensional structures capable of binding their targets with high affinity and specificity, basically mediated by complementary shape interactions (Ellington and Conrad, 1995), (Gold *et al.*, 1995). They can be isolated from vast combinatorial sequence libraries by SELEX (systematic evolution of ligands by exponential enrichment), an *in vitro* selection process (Tuerk and Gold, 1990). The SELEX method has been applied to many different targets ranging from small organic molecules (Famulok, 1999) to large proteins (Gold *et al.*, 1995) and even viruses (Pan *et al.*, 1995) or parasites (Göringer *et al.*, 2003). Moreover, in most cases aptamers not only bind their cognate protein but also efficiently inhibit its function. Thus, aptamers represent an interesting compound class that can be easily obtained and used for assessing the function of a defined protein target.

Also known in the prior art are retroviral expressed peptide libraries containing random sequences (Tolstrup *et al.*, 2001). Retroviral libraries expressing cyclic peptides flanked with dimerization sequences have been successfully used in functional screens of cell cycle inhibitors (Xu *et al.*, 2001). However, this technology has been applied for isolating intracellular peptides and has not resulted in peptidic drugs due to difficulties in delivery as discussed herein.

Another genetic technology for screening bioactive peptides, genetic suppressor element (GSE) methodology, takes advantage of libraries expressing randomly fragmented pieces of cDNAs. While GSE libraries carry natural sequences and are therefore enriched for bioactive clones, they are not adapted to be efficiently or effectively screened for secreted peptides. In spite of the high potential for the discovery of novel drug targets and the development of novel peptide drugs, GSE and random peptide intracellular expression libraries have not had broad application, mainly due to difficulties

in construction, low efficacy, and complicated high-throughput functional screening methodology.

As described previously in this introduction, intrabodies have been shown to be a valuable tool for studying numerous aspects of biological processes, opening up new experimental opportunities to analyze the function of a wide range of cellular molecules. They not only exhibit highly specific molecular recognition properties but are also able to modulate the function of their cognate targets in a highly specific manner.

As intrabodies are derived from the virtually unlimited repertoire of antibodies, there are basically no limitations on the choice of target molecule. With intracellular antibodies, it is possible to develop a reagent that prevents particular associations but spares others. Once an intrabody has been developed, able to inhibit a cellular function, it provides a direct biochemical handle on the recognized protein and its binding partners. Intrabodies have therefore a great added value for functional proteomics and the study of protein–protein interactions, and the development of highly specific inhibitors.

3. Model 1: The regulation of mast cells responses

3.1 Mast cells: Origin and functions

Mast cells are primitive immune cells that appear early in evolution and have since evolved into multifunctional cells in vertebrates. They have been long recognized as initiators of IgE-dependent allergic diseases but it is now realized that they also play a fundamental role in innate and adaptive immune responses to infection as well as inflammatory autoimmune diseases (Galli *et al.*, 1999) (Metz *et al.*, 2008) (Abraham and St John, 2010). In addition, there is evidence that mast cells participate in inflammatory responses to incipient tumors which may either facilitate or retard tumor growth depending on the type of cancer (Galinsky and Nechushtan, 2008) (Ribatti and Crivellato, 2012). Other suspected non-immunological roles for mast cells include promotion of angiogenesis, tissue remodeling, and wound healing (Noli and Miolo, 2001) (Norrby, 2006). As such, mast cells occupy an unstable position in that their responses to endogenous and exogenous stimuli can be detrimental as well as beneficial to the host.

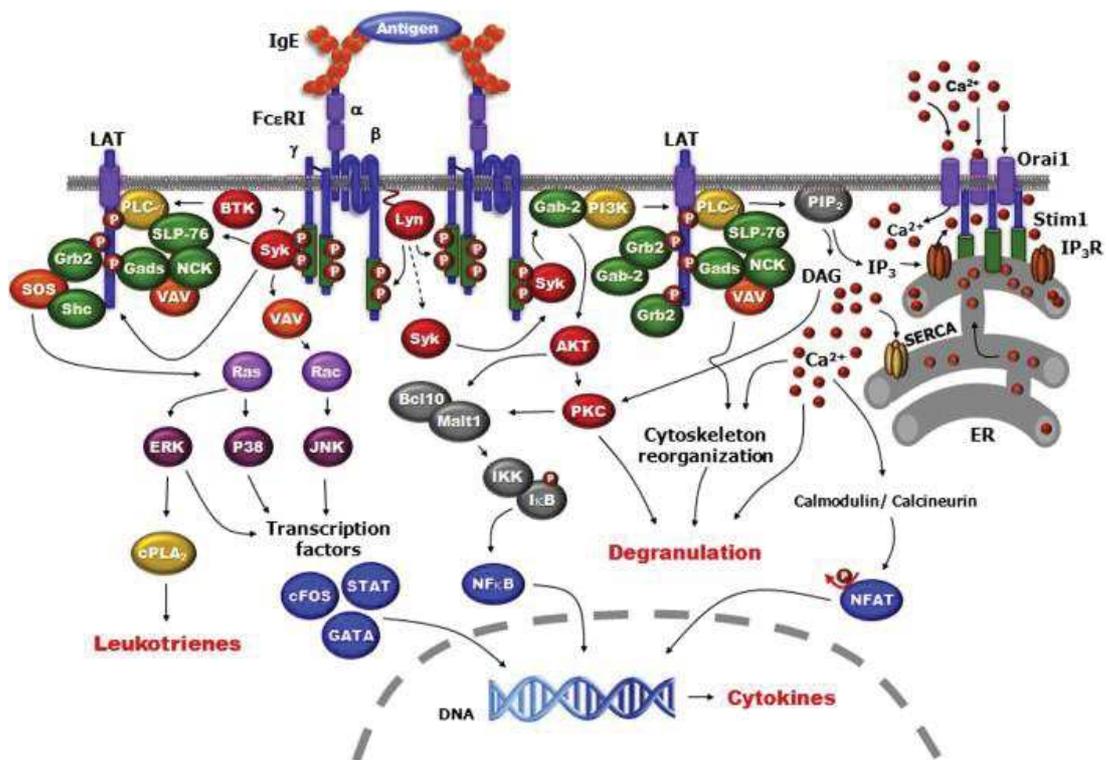


Figure 8: Model depicting major molecules and events in mast cell activation (Siraganian et. al, 2002)

3.2 Signalization pathways

The localization of mast cells in tissues as well as their replication and differentiation into distinct phenotypes are typically determined by the chemical environment of their final tissue destination (Collington *et al.*, 2011). Flexibility in function is also conferred through expression of multiple types of receptors (Galli and Tsai, 2010). These include Fc, cytokine, tyrosine kinase, and trimeric G protein-coupled receptors (Gilfillan and Tkaczyk, 2006). In addition, mast cells can respond to microbial products through pattern recognition receptors (PRR) which include the Toll-like receptors (TLRs) and the recently recognized nucleotide-binding oligomerization domain (NOD)-like receptor (NLRs) (Dawicki and Marshall, 2007). Many of these receptors can act in concert with the high affinity IgE receptor (Fc ϵ RI) to substantially enhance mast cell responses to antigens or alter the pattern of response such that production of cytokines, for example, predominates over degranulation (Qiao *et al.*, 2006). Fc ϵ RI is a member of the immunoreceptor family that includes the B cell receptor, T cell receptor and immunoglobulin receptors. These receptors have intracellular subunits which contain a sequence named the immunoreceptor tyrosine-based activation motif (ITAM) that has two tyrosine residues separated by 6–8 amino acid residues. Fc ϵ RI is a tetramer formed by the complex $\alpha\beta\gamma_2$ chains.

The α -chain binds the Fc portion of IgE at ratio of 1:1 while the β - and γ -chains contain ITAMs in their cytoplasmic domains. Because Fc ϵ RI has no intrinsic enzymatic activity, the activation of non-receptor protein tyrosine kinases is essential for cell activation. Aggregation of Fc ϵ RI results in phosphorylation of the two tyrosine residues within the ITAMs by the Src family of tyrosine kinase Lyn that associate with the receptor. Phosphorylated ITAMs then serve as a docking sites for the tyrosine kinase Syk; the binding of Syk through its SH2 domains results in a conformational change of Syk and its tyrosine phosphorylation due to autophosphorylation and transphosphorylation (Zhang *et al.*, 2000) which increases its enzymatic activity (Figure 8). Once activated, Syk plays a pivotal role in the propagation of downstream signals. The subsequent Syk- and/or Lyn-mediated tyrosine phosphorylation of the transmembrane adaptor molecule LAT (linker for activation of T cells) is crucial for coordination of the downstream signaling pathways that are required for the release of the various pro-inflammatory mediators. Phosphorylation of LAT results in the recruitment of several types of molecules: cytosolic adaptor molecules, guanine-nucleotide-exchange factors and adaptor molecules, and signaling enzymes, such as phospholipase C γ (PLC γ). These interactions with LAT result in the formation of macromolecular signaling complexes, and allow the diversification of downstream signaling that is required for the release of the various pro-inflammatory mediators. For instance, the activation of

PLC γ leads to the production of the second messengers diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP $_3$) which, respectively, result in the activation of protein kinase C (PKC) and the rise in intracellular calcium [Ca $^{2+}$] $_i$. This activates a number of transcription factors needed for *the novo* synthesis of cytokines. The increase in [Ca $^{2+}$] $_i$ plays also an important role in granular fusion and exocytosis of allergic mediators, also called mast cell degranulation. Receptor aggregation also activates and recruits negative regulators which limit the intensity and duration of the positive signals. Negative signals are generated by immunoreceptor tyrosine-based inhibition motif (ITIM) bearing receptors (e.g., Fc γ RII β). The tyrosine residue in the ITIM motif is phosphorylated during the activation steps by Src tyrosine kinases and subsequently recruits inhibitory effectors such as the SH2 domain-containing inositol 5'-phosphatase (SHIP) or the Src homology region 2 domain-containing phosphatase (SHP) via their SH2 domains. In some but not all cases, the negative signals are generated by the co-aggregation of the two receptors.

3.3 Mast cell degranulation

The acute reactions that occur as a result of mast cell activation are initiated as a consequence of degranulation and the generation of lipid-derived mediators, whereas more chronic mast cell-mediated symptoms are an outcome of the delayed generation of chemokines, cytokines and growth factors which follow enhanced gene expression (Metcalf *et al.*, 1997). The process of degranulation occurs within seconds of mast cell activation and the initial rapid phase is essentially complete within 5–10 minutes. Although mast cell proteases such as tryptase, chymase and carboxypeptidase, constitute the major components of mast cell granules, histamine is the predominant granule mediator of acute reactions to mast cell activation. It is localized primarily in mast cells and basophiles although its link to anaphylactic and inflammatory reactions was suspected long before its recognition as a major constituent of mast cells. Histamine is sequestered in mast cell granules by proteoglycans such as heparin and chondroitin E. On release, it readily diffuses through tissues and the circulatory system but does not penetrate the CNS.

3.4 The RBL-2H3 cell line

RBL-2H3 is a basophilic leukemia cell line isolated and cloned from Wistar rat basophilic cells that were maintained as tumors. These cells express the high affinity IgE receptor, Fc ϵ RI. Their activation following Fc ϵ RI aggregation leads to the secretion of early and late allergic mediators. RBL-2H3 cells have been the model for studies of signalization pathways following the stimulation of Fc ϵ RI.

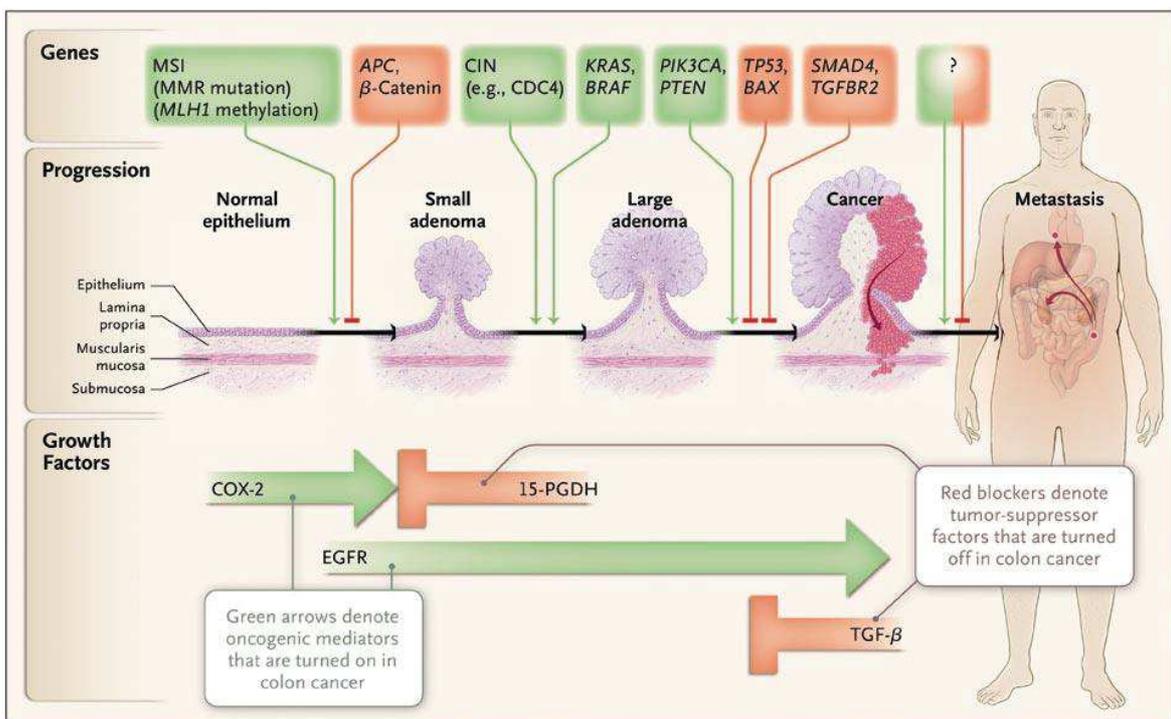


Figure 9: The adenoma–carcinoma sequence model: Molecular events that drive the initiation, promotion, and progression of colorectal cancer (Sanford et al., 2008)

4. Model 2: Colorectal cancer and the resistance to Cetuximab

4.1 Introduction

Colorectal cancer (CRC) is responsible for approximately 608,000 deaths worldwide (8% of all cancer-related deaths), making it the fourth most common cause of cancer-related death. As with the incidence rates, mortality rates are higher among males than females (<http://globocan.iarc.fr>). It is the third most common cancer among males (663,000 cases, 10% of the total) and the second most common cancer among females (570,000 cases, 9.4% of the total) worldwide. There is a variation of 10-fold in the incidence rates for both genders worldwide: the United States, Australia and Western Europe present the highest rates, Africa (excepting Southern Africa) and South-Central Asia have the lowest rates, while rates are intermediate in Latin America. In all areas, incidence rates are higher in men than women. Risk factors for colorectal cancer include increased age, gender (men are more predisposed to the development of colorectal cancer), the presence of inflammatory bowel disease, certain hereditary conditions and a family history of colorectal cancer. It is unlikely that these risk factors will change. However, it should be noted that even in the absence of predisposing factors, the general population has an average risk of developing the disease, and approximately 75% of all cases of colorectal cancer occur in individuals with no known predisposing factors for the disease (Burt *et al.*, 1990).

4.2 The adenoma–carcinoma sequence model

Cancer development is regarded as a multistep process involving tumor initiation, where genomic changes lead to a malignant phenotype, tumor promotion, involving the growth of a mutated cell clone, and tumor proliferation, where more aggressive tumor cells further grow, become invasive and start to metastasize (Lin and Karin, 2007). Concerning these steps, CRC is probably by far the best studied model of tumorigenesis; Fearon and Vogelstein initially proposed a model, later called the adenoma–carcinoma sequence model (Figure 9), in which certain mutations were directly related to distinct stages of tumor development (Fearon and Vogelstein, 1990). In this model, tumor initiation was triggered by mutations of the adenomatous polyposis coli (APC) gene, which allowed adenoma formation and the development of dysplastic crypts. Tumor promotion and progression was supported by a sequence of further mutations including *KRAS*, *SMAD4*, and *p53*, which enabled further adenoma growth, the expansion of individual cell clones, and tumor invasion and metastasis.

Subsequent research showed that mutations observed in CRC are associated with two types of genomic instability and both harbor mutations of different genes. Chromosomal instability (CIN), which includes the presence of various structural or numerical chromosomal changes, can be found in about 65–70% of CRCs and has been related to mutations of the adenoma–carcinoma sequence model (Miyazaki *et al.*, 1999). In contrast, 15% of CRCs develop microsatellite instability (MSI) as a result of defective DNA mismatch repair (MMR) genes. MSI is characterized by single nucleotide mutations and variations in the length of microsatellite sequences. It has been regarded as a positive predictive marker in patients with CRC, since patients with MSI + CRC have an improved prognosis in comparison to CIN + CRC. Interestingly, MSI is only irregularly associated with the above mentioned mutations of CIN + CRCs. Besides defects in MMR genes, MSI has been related to mutations of transforming growth factor β receptor 2 (TGF β R2), insulin-like growth factor 2 receptor (IGF2R), BAX, and others (Walther *et al.*, 2009). Despite these differences in CIN + and MSI + CRCs, the observed mutations are associated with similar or even identical molecular pathways that are critical for the individual steps of tumor development.

Approximately half of all protein-encoding genes in the human genome contain CG-rich regions in their promoters or CpG islands. Aberrant DNA methylation, in the form of hypermethylation of CpG islands, results in repression of transcription in tumor suppressor genes. For example, inactivation of the mismatch repair gene *MLH1* by promoter methylation is the molecular basis for microsatellite instability in sporadic microsatellite unstable colon cancers (Samowitz, 2007). This phenomenon of tumor alteration via epigenetic silencing associated with dense hypermethylation of CpG islands, and their complex interplay with modifications in histone structure, provides an alternate mechanism to genetic inactivation of tumor suppressor genes via loss or mutation (Esteller, 2008). The presence of widespread CpG island methylation in a tumor is termed the CpG island methylator phenotype, or CIMP, on which several studies has been performed (Marisa *et al.*, 2013) (de Melo *et al.*, 2014).

4.3 Current therapies for colorectal cancer

Primary prevention of colorectal cancer is based on the implementation of screening tests to detect and remove pre-cancerous lesions or to discover and treat cancer at its earliest stages. The survival rate of patients with colorectal cancer depends on the clinical and pathological stage of the disease at diagnosis. Patients with cancer limited to the bowel wall at diagnosis have a 5-year survival rate of 90% (Mandel *et al.*, 1993). The survival rate is reduced to 35-60% if the lymph nodes are involved, and drops to less than 10% when the disease is metastatic (Wingo *et al.*, 1995).

There is full agreement among health care professionals regarding the essential role of surgery in the treatment of colorectal cancer. Further improvements are obtained through the use of treatments combining surgery, chemotherapy and radiotherapy, and in particular adjuvant and neoadjuvant therapies.

Various combinations of chemotherapeutic drugs are currently employed. 5FU, which was the first chemotherapeutic agent used to treat colorectal cancer, is still used today or in combination, as with FOLFIRI (folinic acid, FU and irinotecan) or FOLFOX (folinic acid, FU and oxaliplatin) protocols. New biological compounds, such as Bevacizumab, Cetuximab and Panitumumab, are also used in the treatment of colorectal cancer. In particular, vascular endothelial growth factor (VEGF) has proven to be the most powerful angiogenic factor. VEGF binds to its receptor, activating several intracellular signal transduction pathways (Tol and Punt, 2010). In addition, VEGF receptor-2 was expressed on the surface of colorectal cancer cells in approximately 50% of samples analyzed (Duff *et al.*, 2006). The role of VEGF in cancer progression is still open to discussion but we know that VEGF and other pro-angiogenic factors, in addition to their traditional role in neo-vascularization, also stimulate the degradation of the extracellular matrix, leading to the proliferation and migration of endothelial cells (Dvorak, 2002). For this reason, both VEGF and VEGF receptors are valuable therapeutic targets. Bevacizumab is a humanized monoclonal antibody against VEGF. It binds to VEGF, thereby preventing VEGF from binding to its receptor and halting the subsequent activation of intracellular signal transduction pathways. Previously, it was thought that high microvessel density and VEGF expression represented negative prognostic markers in patients with metastatic colorectal cancer (Des Guetz *et al.*, 2006). However, it now appears that the response to Bevacizumab is independent of VEGF expression or high microvessel density (Jubb *et al.*, 2006). Predictive markers of response to Bevacizumab have not yet been identified.

The epidermal growth factor receptor (EGFR) is a member of the ErbB transmembrane tyrosine kinase receptor family. Various ligands (EGF, TGF- α , epiregulin and amphiregulin) bind to this receptor and stimulate several intracellular signal transduction pathways (RAS/RAF/MAPK, PI3K/Akt and JAK-STAT), leading to cell proliferation, differentiation and the arrest of apoptosis (Scaltriti and Baselga, 2006). In addition, EGFR has been observed to be overexpressed on the surface of colorectal cancer cells in approximately 40-70% cases, and has been associated with a decrease in survival (Spano *et al.*, 2005). Currently, two new anti-EGFR molecules are used to treat colorectal cancer: the monoclonal antibodies Cetuximab and Panitumumab. Cetuximab (Erbbitux[®]) is a chimeric IgG1 antibody and Panitumumab is a fully human IgG2 antibody. Notably, it was determined that

Cetuximab can only be used in the absence of *KRAS* oncogene mutations (De Roock *et al.*, 2008).

KRAS mutations are observed in approximately 40% of colorectal cancer patients (Andreyev *et al.*, 1998), and cause the constitutive activation of the RAS/RAF/MAPK pathway, which is not linked to EGFR activation by ligand binding (Benvenuti *et al.*, 2007). Since patients without *KRAS* wild-type cancer cannot benefit from treatment with anti-EGFR antibodies, the European Medicines Agency has approved the use of these antibodies only in patients with *KRAS* wild-type tumors (<http://www.ema.europa.eu/ema/>).

Among patients with *KRAS* wild-type tumors, anti-EGFR therapy appears to be further limited to patients with *BRAF* wild-type tumors (Tol *et al.*, 2009), although a recent study found that the *BRAF* mutation was a prognostic rather than a predictive marker (Bibeau *et al.*, 2009). Anti-VEGF and anti-EGFR antibodies have a proven efficacy. At present, Bevacizumab, Cetuximab and Panitumumab are the standard therapeutic options for metastatic colorectal cancer. Bevacizumab is administered in combination with chemotherapy as a first-line treatment, while anti-EGFR antibodies appear to be more beneficial in later-line treatment. Furthermore, anti-VEGF and anti-EGFR antibodies, in combination with chemotherapeutic agents, increased the life expectancy of patients with metastatic colorectal cancer by approximately 2 years.

The combination of anti-VEGF and anti-EGFR therapy has been proposed, since VEGF has many intracellular pathways in common with EGFR, and because colorectal cancer cells and endothelial cells express both EGFR and VEGF receptor-2 (Duff *et al.*, 2006). Although pre-clinical studies demonstrated that EGFR ligands up-regulated VEGF expression (Niu *et al.*, 2002) and that the combination of anti-angiogenic and anti-EGFR agents might have significant antitumor activity (Jung *et al.*, 2002), numerous clinical trials have found that not only is the combination of the two antibodies less effective than the individual agents combined with other chemotherapeutic drugs, but also that their combination results in a reduction of progression-free and overall survival medians (Tol and Punt, 2010). These findings led to a requirement for *KRAS* mutational testing of metastatic CRC (mCRC) and restricting the use of the anti-EGFR antibodies to patients with *KRAS* wild-type mCRCs (Heinemann *et al.*, 2009). On its face this represents another advance toward effective personalized cancer treatment (Heinemann *et al.*, 2013).

Yet, the increases in progression-free survival for patients with *KRAS* wild-type tumors are modest (Amado *et al.*, 2008), half of patients with *KRAS* wild-type tumors fail to respond (Pentheroudakis *et al.*, 2013), and more importantly, eventually all patients develop resistance to treatment and

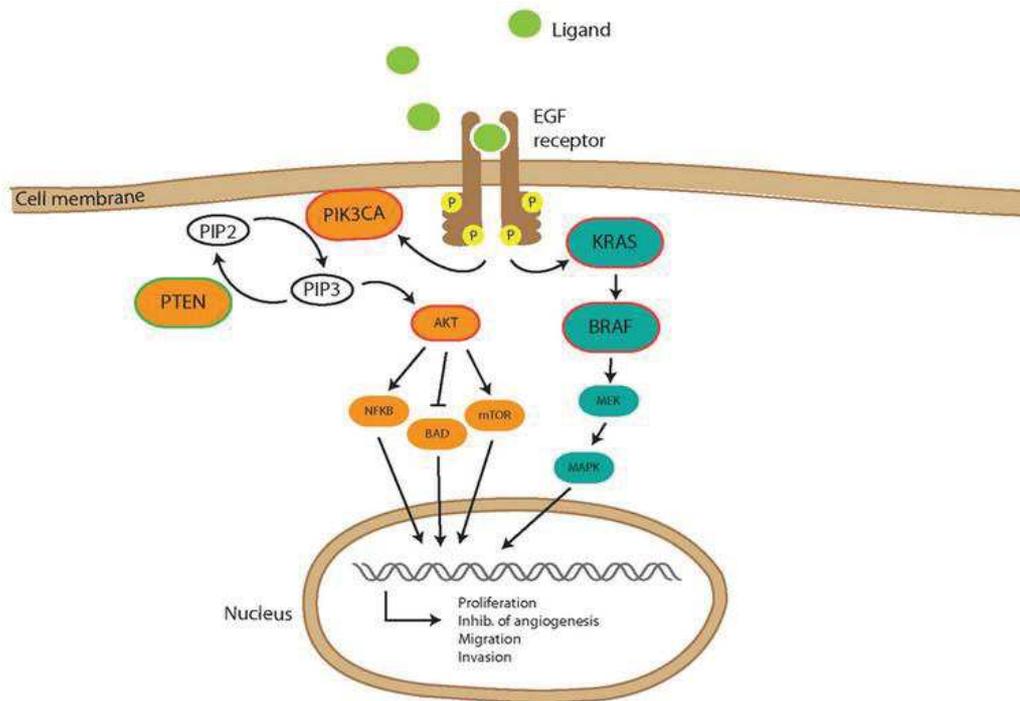


Figure 10: EGFR pathway and signaling through the MAP kinase pathway (Berg and Soreide, 2012).

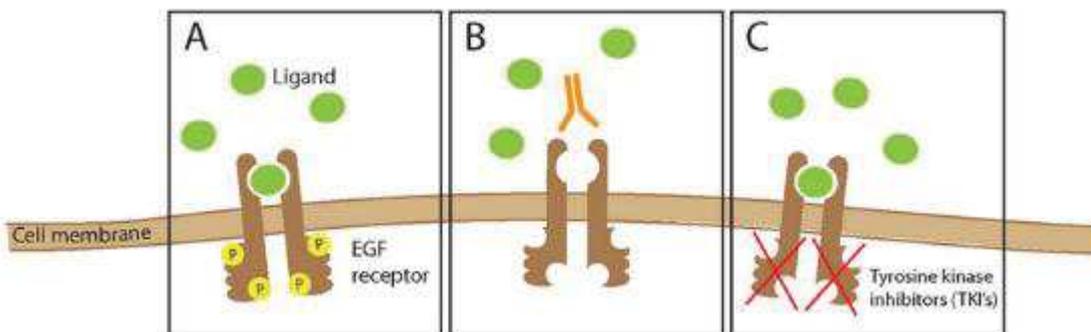


Figure 11: A. The EGFR is activated by ligand binding to the extracellular domain, which results in activation B. Monoclonal antibodies hinder binding of ligands. C. Small molecule tyrosine kinase inhibitors prohibit EGFR phosphorylation (Berg and Soreide, 2012).

relapse (Misale *et al.*, 2012). It can be concluded that few colon tumors are actually *KRAS* wild-type and that most, if not all, colon tumors carry abnormal levels of *KRAS* mutation. Furthermore, the use of EGFR-targeted therapies to treat predominantly *KRAS* wild-type tumors creates an opportunity for the outgrowth of *KRAS* mutant subpopulations, which can lead to acquired resistance to treatment and relapse. The significance of these findings will be discussed in terms of their broader implications for developing effective personalized cancer treatments.

4.4 The EGFR pathway and its implication in cancer therapy

The EGFR is a member of the family of receptor tyrosine kinases, and the gene is considered an oncogene. EGFR activation results in signaling through the MAP kinase pathway, and through the PI3K/AKT pathway, causing cellular growth and progression, proliferation, angiogenesis, and invasion (Figure 10). EGFR expression is reported to be elevated in 60-80% of colorectal tumors, investigated using immunohistochemistry (IHC) (Goldstein and Armin, 2001) but is generally not considered a good predictive marker (Yang *et al.*, 2012).

The employment of anti-EGFR antibodies represents a backbone of the therapeutic options for mCRC with relevant efficacy in both first- and second-line treatment regimens. However, this therapy is poorly effective or ineffective in unselected patients. Mutations in *KRAS*, *BRAF*, and *PIK3CA* genes have recently emerged as the best predictive factors of low/absent response to EGFR-targeted therapy. As the increased knowledge of tumor heterogeneity and genetic alterations progresses, it exemplifies the need for further personalized medicine in modern cancer management.

4.5 Treatment directed towards the Epidermal Growth Factor Receptor

Epidermal growth factor receptor signaling is a response of binding of cytokines, hormones, or growth factors to the extracellular domain of the receptor (Figure 11A). The monoclonal antibody Cetuximab selectively binds to the extracellular domain of EGFR and sterically blocks the access of growth factors to a key ligand binding region on domain III of the receptor. The ligand binding site on domain I is unaffected, but the growth factor must engage sites on both domains I and III for high-affinity binding that activates the receptor, so efficient blockade of either one is sufficient to debilitate the receptor (Figure 12).

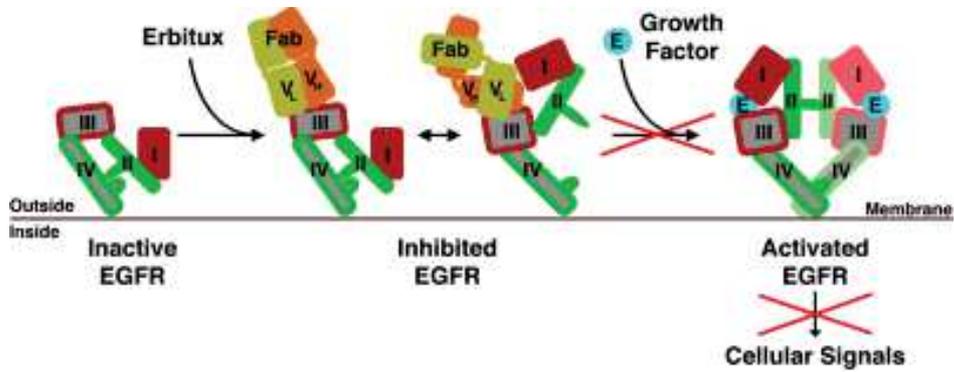


Figure 12: Model for ligand-induced dimerization by Cetuximab and its blockade by Cetuximab. The domains of the extracellular region of EGFR are shown; with domain I in red, domain II in green, domain III in gray with red border, domain IV in gray with green border, and ligand in cyan (Li et al., 2005).

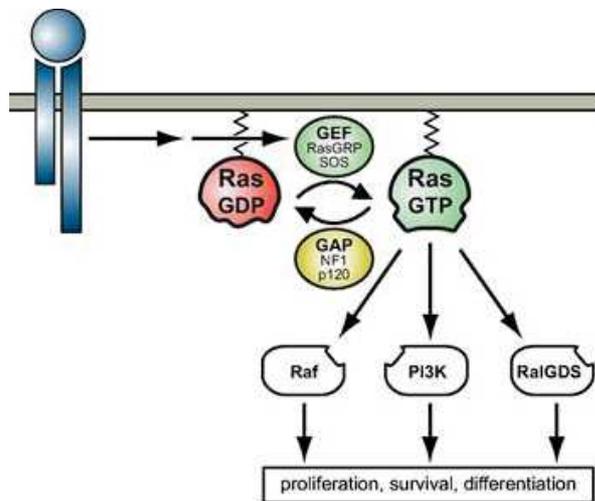


Figure 13: The Ras switch (Ward et al., 2012).

The blockade of receptor dimerization by Cetuximab disables EGFR autophosphorylation, downstream signaling, and consequently cell growth and proliferation terminates (Figure 11B) (Citri and Yarden, 2006). This is believed to be the critical factor in the observed antitumor effects of this antibody *in vivo* (Mendelsohn and Baselga, 2003).

Both Cetuximab and Panitumumab mAbs are used as treatment for mCRC. However, as mentioned before, treatment with mAbs has been found to be efficient in only a small percentage of patients. This implies the presence (or absence) of certain mutations, as found in *KRAS*.

4.6 The *KRAS* mutational status

KRAS is an oncogene encoded on the short arm of chromosome 12. RAS proteins are central signal transduction molecules, which act as molecular switches through cycling between an active, GTP-bound and an inactive, GDP-bound state (Figure 13) (Ward *et al.*, 2012). In its GTP-bound form, RAS interacts with and regulates a vast spectrum of functionally diverse downstream effectors. Many mutations of RAS affect the on/off balance and RAS mutations at codon 12, 13 and 61 lead to constitutive active RAS protein and the subsequent hyper-activation of the MAPK pathway.

The gene is mutated in ~40% of CRCs, mainly in the codons 12 and 13, and to a lesser degree in codons 61 and 146 (Forbes *et al.*, 2008). The most frequent mutation is substitution of glycine 12 to serine and is found in approximately 80% of cases. It has been shown that mutations in codons 12, 13, and 61 all predict a lack of response to Cetuximab. It is suggested that mutations in both codon 13 and 146 do not affect Cetuximab efficacy (De Roock *et al.*, 2011). To date, *KRAS* is the only molecular marker used to select patients eligible for mAbs treatment. If relying on test results from the primary tumor only, patients with *KRAS* mutation would not be offered anti-EGFR therapy, even if metastatic lesions without the mutation could benefit from such treatment. A focus on accordance between the genotype in the primary tumor and metastases from the same patient has gained attention recently, as mutation status in the primary tumor might not reflect the mutational situation in metastases of the patient (Cejas *et al.*, 2012). Discrepancies in mutational status within the same individual are of importance, e.g., when using treatment targeting the EGF receptor, as *KRAS* status is used in clinical practice to decide the use of anti-EGFR treatment. *KRAS* mutations are considered to be an early event in colorectal tumorigenesis, and consequently a high degree of correspondence in mutational status between tumor cells of the same individual is assumed. Discordance in *KRAS* genotype between primary tumor and metastasis might be explained by dissemination of cells at an early step

in tumor development, heterogeneity in mutation status within the primary tumor, or the fact that the metastasis originates from an undetected second primary. Loss of *KRAS* mutations during disease progression, or tumor heterogeneity, giving rise to different subclones of metastases, is suggested to be causative for discordance between primary and metastatic tumors. A lack of response in *KRAS* wild type patients have been suggested to be caused by mutations in one or several metastases.

Mutations in *KRAS* and the downstream effector *BRAF* are thought to be mutually exclusive (Garnett and Marais, 2004), although some overlap has been reported (Seth *et al.*, 2009). Activating mutations in *BRAF* are reported in ~10% of CRC samples, all observed within the kinase domain of the gene, substitution of valine by glutamate in position 600 of the protein being the most prevalent (Garnett and Marais, 2004). As for *KRAS*, mutations in *BRAF* predict a lack of response to anti-EGFR therapy as well as poor prognosis (Barault *et al.*, 2008).

4.7 Other pathways activated by EGFR

In addition to signaling downstream through *BRAF*, EGFR also activates the AKT pathway (Figure 10). The phosphatase PTEN and the phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit (PIK3CA) have opposing roles in this pathway, converting PIP3 (phosphatidylinositol 3-phosphate) to PIP2 (phosphatidylinositol 2-phosphate). Elevated levels of PIP3 results in hyperphosphorylation of AKT, either due to loss of PTEN activity or activating mutations in PIK3CA. Hyperphosphorylated AKT inhibits the cell from undergoing apoptosis. In patients with metastatic CRC that received Cetuximab treatment, the presence of a *KRAS* wild type with PTEN loss, or having activating mutations in *PIK3CA*, was reported to predict no response and reduced overall survival (Laurent-Puig *et al.*, 2009).

4.8 Treatment directed towards targets downstream of EGFR

There are several compounds in clinical development which selectively inhibit mutated *BRAF* (Cichowski and Jänne, 2010). But these agents should be avoided in *KRAS*-mutated cancers, as these will have increased activity through *BRAF* (Heidorn *et al.*, 2010). Further downstream the EGFR signaling cascade from *BRAF* is the mitogen-activated protein kinase (MEK). MEK inhibitors have been suggested for patients being mutant for either *KRAS* or *BRAF* (Duffy and Kummar, 2009). Results have shown that *BRAF* mutant cell lines respond to MEK inhibitors, but low response rates and ocular toxicities in clinical trials have obstructed the investigation (Lemecch and Arkenau, 2012). *BRAF* mutated cell lines were found to be more sensitive to MEK inhibitors than *KRAS* mutated cell lines, probably because of the fact that *KRAS* has several downstream effectors (Chappell *et al.*, 2011).

However, *BRAF* gene amplification has been shown to result in resistance to both MEK and BRAF inhibitory treatment.

4.9 Induction of mutations during anti-EGFR treatment

The development of therapy-related resistance during anti-EGFR antibody treatment in mCRC has been investigated in the light of treatment-induced shift in mutation status for *KRAS* and *BRAF*. *KRAS* and *BRAF* mutation status was shown to be highly concordant in primary tumors before and after anti-EGFR therapy, indicating that therapy-induced resistance is likely not induced by mutations in the hotspot regions of these genes (Gattenlöhner *et al.*, 2009). However, more recently, a study found that 35% (of 24 patients) whose tumors were initially *KRAS* wild type developed detectable mutations in *KRAS* in the serum, three of which developed multiple different *KRAS* mutations (Diaz *et al.*, 2012). This supports the theory that the emergence of *KRAS* mutations is a mediator of acquired resistance to EGFR blockade and that these mutations can be detected in a non-invasive manner (by blood test). Further, it may explain why solid tumors develop resistance to targeted therapies in a highly reproducible fashion. Recently, an acquired EGFR ectodomain mutation (S492R) was described that prevents Cetuximab binding and confers resistance. Patients with this mutation, however, retain binding to and are growth inhibited by Panitumumab (Montagut *et al.*, 2012). Another study found frequent coexistence of *KRAS* and *PIK3CA* mutations, pointing to the need for targeting both pathways to overcome resistance (Garrido-Laguna *et al.*, 2012).

Our laboratory has a strong expertise in the field of the intracellular immunization and its application for the identification of new therapeutic targets and new drugs. Previous works performed in our group have shown the successful employment of this approach for the inhibition of protein-protein interactions of the tyrosine kinase Syk and for the discovery of non-enzymatic inhibitors of Syk and allergy (Mazuc *et al.*, 2008) (Villoutreix *et al.*, 2011). The works of our group also demonstrated that intrabodies can specifically target a post-translational modification of a protein within a living cell (Cassimeris *et al.*, 2013). This is particularly important since it demonstrates one of the main advantages of intrabodies compared to RNAi.

More recently, we have performed a phenotypic screening assay in a model of allergy which led to the identification of a new therapeutic target implicated in mast cell activation. This work was protected by the European Patent EP 13 305 003.9 “Intracellular phenotypic screening” in 2013 and was published recently (see article Mazuc *et al.*, 2014). My contribution to this work (Part 1) is described in “1.4 Characterization of inhibitory intrabody families”. The goal of my PhD project was to perform a similar intrabody-based screen in cancer cells. For this purpose, I first optimized the scFv library for its intracellular expression and stability under strong reducing conditions (Part 2). Then I used this newly synthesized library for a phenotypic screen in *KRAS*-mutated colorectal cancer cells resistant to Cetuximab (Part 3). My aim was to identify intrabody fragments able to revert the resistant phenotype to a sensitive one, and subsequently identify their potential cellular targets.

1. Phenotypic targeting of mast cells and their role in allergy

Screening for new therapeutic targets and specific inhibitors remains the key to develop new medicines. The majority of methods first identifies a suitable target before designing chemical inhibitors, or identifies inhibitors in a cellular system then look for the drug target. In both cases there is no guarantee of success since some targets have proven to be unsuitable for specific inhibitor design and drugs may simultaneously act on multiple targets and pathways. Targets are frequently identified using small scale *in vitro* and *in vivo* experiments coupled with bibliographic analysis. This has been expanded in recent years to genome-wide screenings, based on mutagenesis or RNAi that have revealed new targets and original regulatory networks.

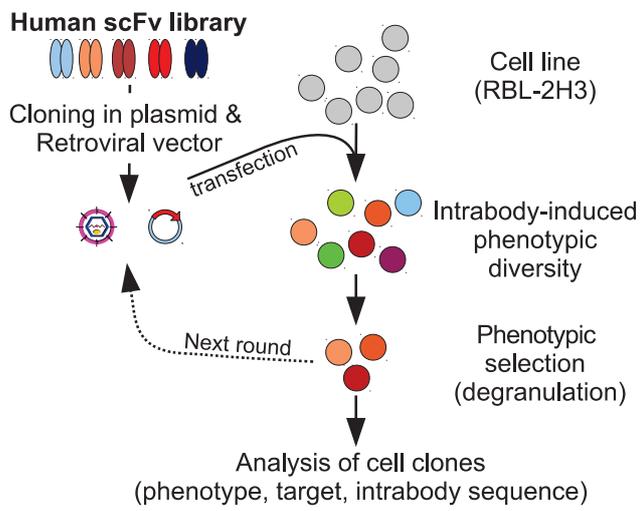


Figure 14: Schematic view of the selection method (see article Mazuc et al., 2014).

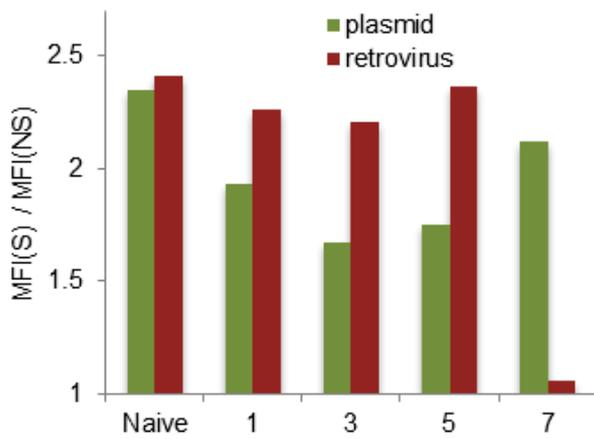


Figure 15: Annexin-V staining of cell populations from the library selection rounds (see article Mazuc et al., 2014)

In all cases, there is however no insurance that a good inhibitor of the target could be developed and translation of a target to a chemical or biological active compound may require years of efforts and eventually fail. The problem can be reversed by direct screening of chemical libraries in a suitable cell-based assay. Several methods to help in the identification of the target have been described but all are long, tedious and not always successful. They may rely on biochemical methods, genetic analyses or computational approaches to not only identify the target but also the secondary proteins that may lead to off-target effects.

1.1 Phenotypic screen and target identification

Here, I describe shortly the results of the intrabody-based phenotypic screen, which strategy is depicted in figure 14. Because the method is based on the intrabody-target interaction, it results in the co-selection of a target with its companion intrabody. The aim is to identify new therapeutic targets in the field of allergy and inflammation. The model, FcεRI-mediated mast cell degranulation of RBL-2H3 basophilic cell line, is visualized through the staining of the exocytotic granules with Annexin V. The IgE-dependent stimulation of the cells leads to the exposure of exocytosing granules and phosphatidylserines that can be monitored, in proportion to the extent of allergic mediator release, by the binding of exogenously added Annexin-V. We used this method to isolate by flow cytometry cell sorting (FACS) the population of intrabody-containing RBL-2H3 cells that displayed an impaired degranulation.

First, an scFv library optimized for intracellular expression was sub-cloned in plasmid and retroviral eukaryotic expression vectors and two libraries of a diversity of 10^9 and $2 \cdot 10^8$ were generated respectively. The recombinant vector pools were subsequently used to transfect the RBL-2H3 mast cell line in order to generate two distinct populations of $5 \cdot 10^7$ transformed cells. Seven rounds of selection were performed, which allowed the enrichment for cells containing intrabodies able to block cell degranulation. This procedure allowed the enrichment for cells containing intrabodies able to block cell degranulation as shown by the decrease of the Annexin V staining of stimulated versus unstimulated cells (Figure 15).

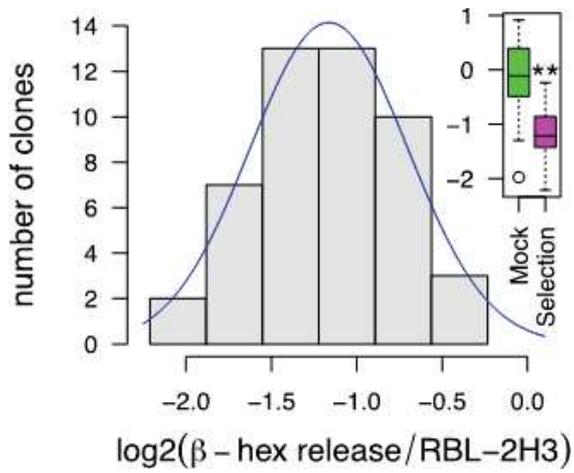


Figure 16: Distribution of the β -hexosaminidase release measured on 48 retroviral clones (see article Mazuc et al., 2014) **: $p < 0.01$ (Student t-test).

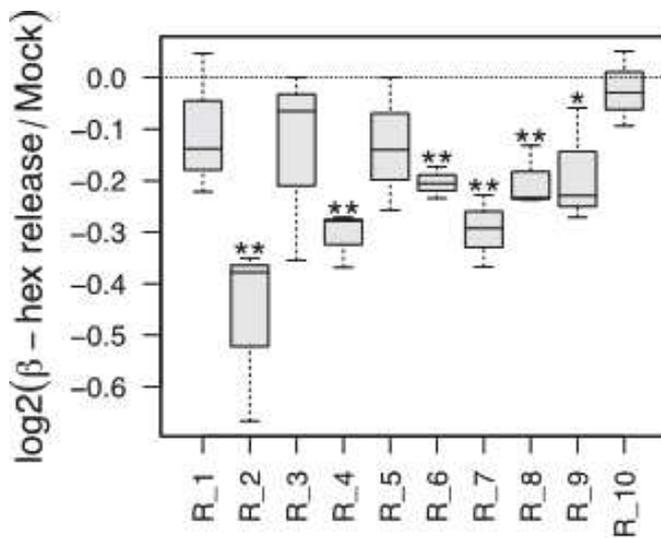


Figure 17: Measure of β -hexosaminidase release by retroviral infected populations (see article Mazuc et al, 2014) *: $p < 0.05$; **: $p < 0.01$; (Student t-test).

1.2 *Plasmid library sequences*

To identify individual intrabody sequences responsible for the inhibitory phenotype, we generated stable RBL-2H3 clones from the pool of plasmids obtained from the seventh round of selection. One hundred and twenty-six stable clones were tested in a degranulation assay based on the measurement of the FcεRI-mediated release of the enzyme β-hexosaminidase. The low-degranulating clones that express inhibitory intrabodies represented about 20% of the clones. The analysis of the intrabody sequences expressed in 36 clones revealed a high diversity with 1-2 different sequences expressed in each clone.

1.3 *Retroviral library sequences*

An aliquot of cells from selection round 7 was seeded at limiting dilution and 48 isolated clones were analyzed in triplicate by measuring the release of β-hexosaminidase (Figure 16). Here, a significant inhibition of the β-hexosaminidase release was observed (54% on average). This showed that the retroviral library selection was more powerful than the plasmid one, presumably because of the presence of fewer intrabodies per cell, which reduced the co-selection of passenger intrabodies to a minimal. In order to analyze the intrabody diversity evolution in the cell population during the course of selection with the retroviral library, high-throughput sequencing was performed.

Since the VH domain is known to be the most important determinant of antibody affinity and specificity (Ward *et al.*, 1989) only the CDR3 sequences that correspond to the variable part of the VH domains in the library design were analyzed. 2568 VH DNA sequences appeared to be enriched and none significantly depleted during the course of the selection. Enriched sequences were translated and we retained the 108 VH sequences that continuously increased during the course of the selection. These 108 VH represented 40% of the sequences present in the final selected library.

1.4 *Characterization of inhibitory intrabody families*

My contribution in this project was to further evaluate the inhibitory phenotype of 10 families of intrabodies that were retrieved during the last round of selection. Between the naive library and the final selected library, each of these 10 families were enriched about 150-fold, however some individual sequences were enriched more than 500-fold. In addition, whereas the frequency of most of the clones present in the initial library decreased at an exponential rate during the course of the selection, the 10 identified families were strongly enriched.

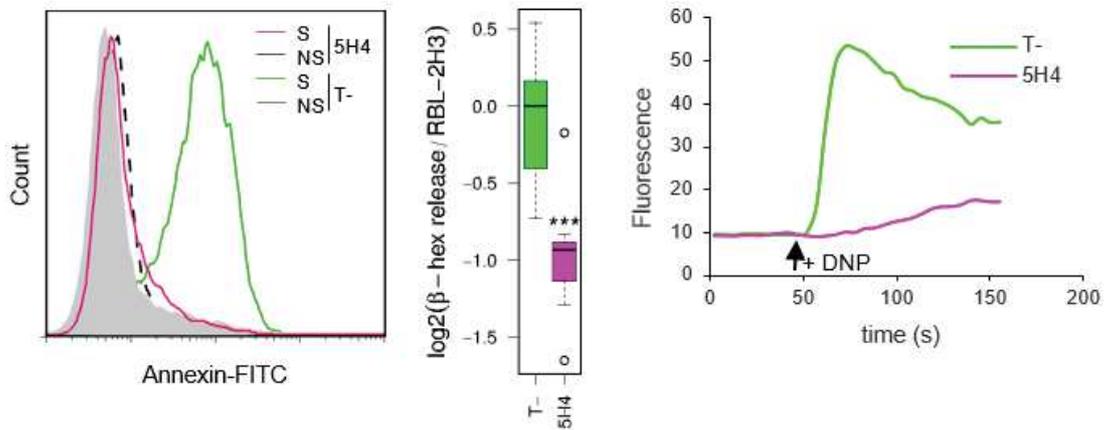


Figure 18: Analysis of stable plasmid clone 5H4. **Left:** Annexin-V staining; **Middle:** β -hexosaminidase release (7 independent replicates); **Right:** measurement of calcium flux. T-: Irrelevant intrabody. S: IgE/DNP stimulated. NS: unstimulated (see article Mazuc et al., 2014).

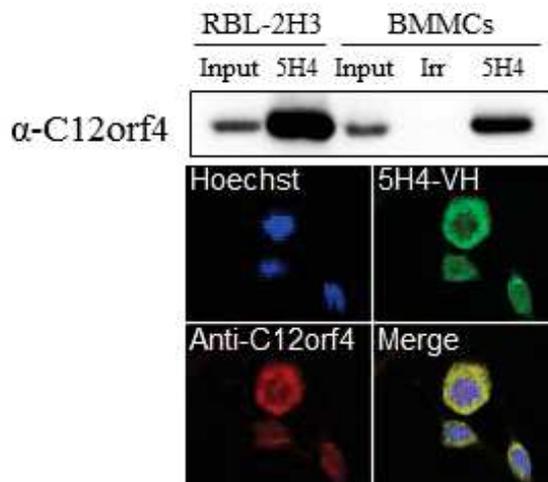


Figure 19: Specific binding of 5H4-VH to C12orf4. **Top panel:** pull-down assay using 5H4-VH, and detection with a commercial anti-C12orf4 antibody. **Low panel:** subcellular localization of C12orf4 analyzed by confocal laser microscopy. **Top left:** Hoechst; **top right:** 5H4-VHFc fusion; **bottom left:** anti-C12orf4 commercial antibody; **bottom right:** merge (see article Mazuc et al., 2014).

To analyze their inhibitory properties, we retrieved the most frequent sequence of each family (R_1 to R_10) (Annex 2). Therefore, we designed an oligonucleotide based on the VH CDR3, amplified the VH and VL sequences, assembled them in a full scFv, and then recloned the reconstituted gene in a retroviral vector. Two independent tests were performed to analyze the inhibitory effect of the intrabodies on FcεRI-induced mast cell activation. The measurement of the secretion of β-hexosaminidase, one of the early allergic mediators released by exocytose, and the release of TNF-α, a newly synthesized cytokine that stimulates inflammation. Both cellular tests were performed in quadruplicate during four weeks. Figure 17 show a representation of the secretion of β-hexosaminidase for the 10 clones in comparison with cells transfected with an empty vector (mock) to which we have calculated the percentages of released mediator. Six of ten intrabodies induced a significant inhibition of degranulation following FcεRI stimulation.

Next, 136 unique VH sequences obtained from 178 clones of the plasmid library (stable clones and randomly picked sequences from the last round of selection) were compared with the 108 VH sequences enriched during the retroviral selection. Only one sequence was found common to both selections, and that was the VH domain of plasmid clone 5H4. As we demonstrated previously, the 5H4 family (R_8) is part of the 10 best families selected from the retroviral sequence analysis and the 5H4-VH sequence was enriched 170-fold during the retroviral selection. Plasmid-derived stable RBL-2H3 clone 5H4 did not show any Annexin-V staining following FcεRI stimulation and both β-hexosaminidase release and calcium flux were strongly inhibited (Figure 18). These results confirmed the inhibitory potential of the retroviral clone R_8 expressing the intrabody 5H4-VH (Figure 17).

Analysis of 5H4 sequence revealed that the gene encoding scFv 5H4 was truncated at its C-terminus because of the presence of a stop codon in the first codon of the light chain CDR3. Since the original scFv library contained only variable CDR3 loops, we reasoned that the inhibitory phenotype of 5H4 was carried by its VH portion. Indeed, isolated human VH domains have already been shown to be efficient intrabodies, particularly when the VH sequence belongs to the human VH3 class as it is the case here.

1.5 Target identification

One of the advantages of phenotypic screening is that intrabodies have high affinity and specificity, which can be used for the identification of their target. In order to identify the cellular target of intrabody 5H4, the gene encoding 5H4-VH was expressed in *E. coli* cytoplasm, and the purified intrabody fragment was used to capture its target from RBL-2H3 extracts. The captured proteins were

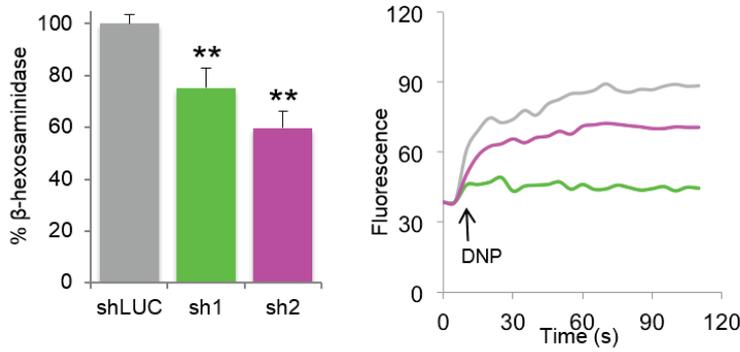


Figure 20: shRNA-induced down-regulation of *C12orf4*. Analysis of β -hexosaminidase release (**left**) and calcium flux (**right**) were performed with cell populations 15 days post-infection (see article Mazuc et. al, 2014).

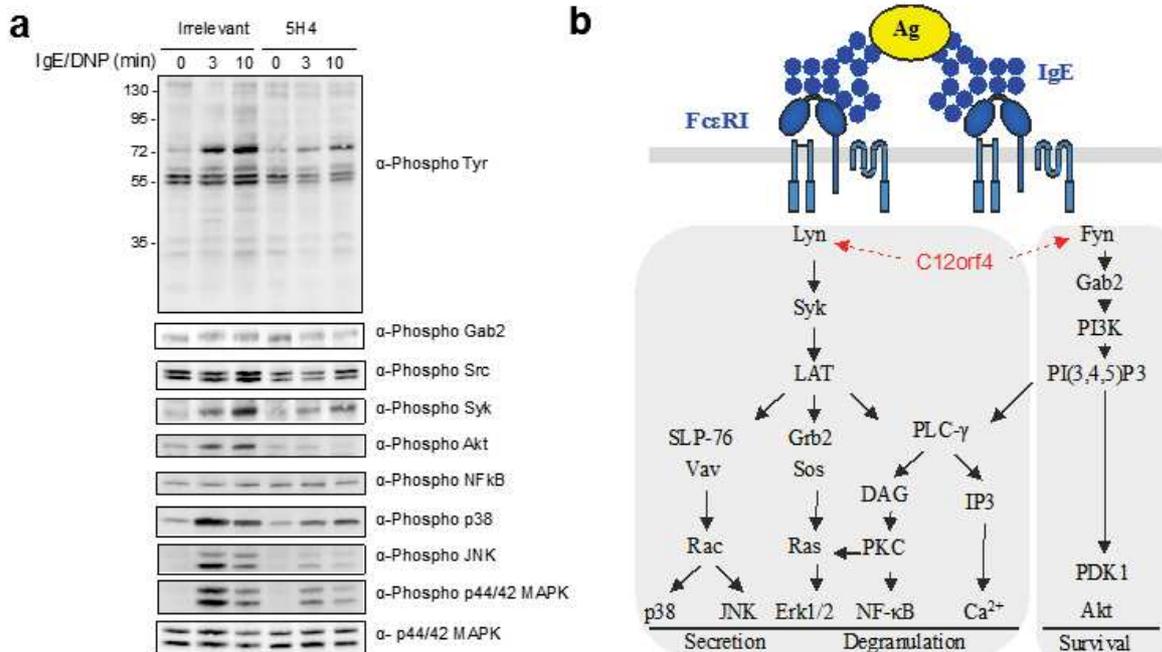


Figure 21: **a)** Western blot analysis of the Fc ϵ RI-mediated phosphorylation of major proteins implicated in mast cell activation. **b)** Schematic view of mast cell signaling pathways (see article Mazuc et. al, 2014).

analyzed by mass spectrometry using an irrelevant VH fragment as a control. A unique protein called LOC297607/C12orf4 was identified in a mast cell line and bone marrow-derived mast cells using a pull-down assay (Figure 19, top panel).

In addition, fluorescence microscopy showed that C12orf4 is a cytoplasmic protein (Figure 19, low panel). Database interrogation revealed that C12orf4 is a protein of unknown function and widely conserved from nematodes to humans.

The lab further characterized the role of C12orf4 in FcεRI-mediated mast cell responses. Short hairpin RNA (shRNA) approach was used for down regulation and modulation of C12orf4 expression in RBL-3H2 cells. The analysis of the degranulation of shC12orf4 transfected cells showed a decrease in β-hexosaminidase release that correlated with an inhibited calcium flux (Figure 20).

In mast cells and basophils, the engagement of FcεRI initiates the activation of the Src kinases Lyn and Fyn and the Syk tyrosine kinase, which allows signal propagation through the phosphorylation and activation of several downstream proteins (Kalesnikoff and Galli, 2008). As such, the impact of targeting C12orf4 on FcεRI-mediated signaling events was investigated and an impairment of both Lyn- and Fyn-dependent signals was noticed. These results are consistent with the defect in the degranulation events (Figure 21). Taken together, our results suggest that C12orf4 plays a role in the early signaling events leading to degranulation.

In summary, this study demonstrates that the selection of intrabodies as phenotype-modulators can be accomplished in cell-based assays using unbiased combinatorial libraries (not preselected by phage panning) and without any previous knowledge of a target.

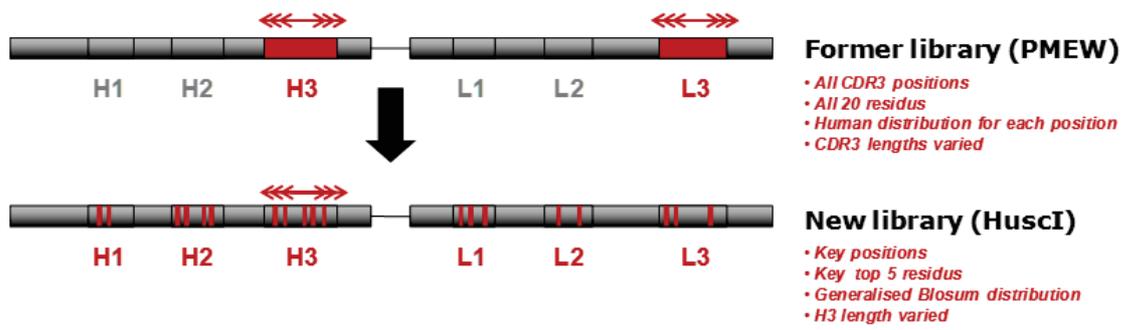


Figure 22: Design scheme of the former (PMEW) versus the new (HuscI) scFv library.

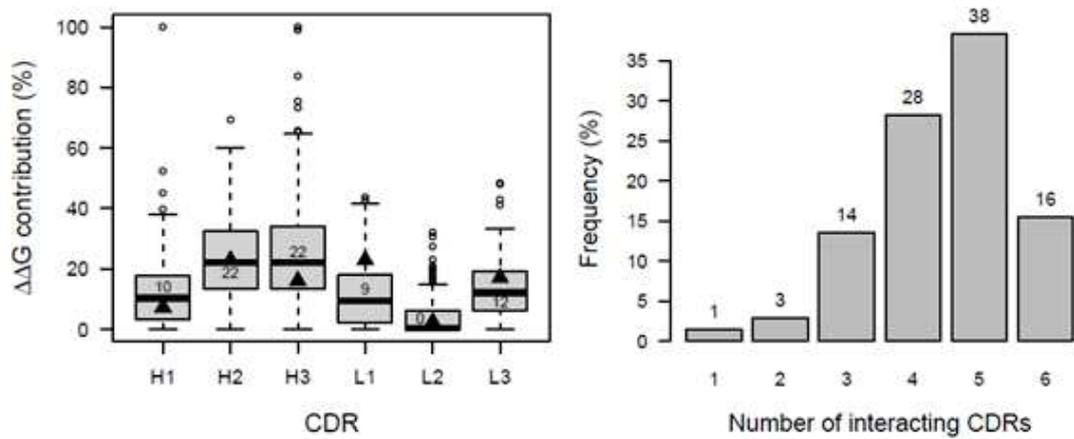


Figure 23: CDR contributions and usage. **Left;** Distribution of the CDR binding free energy contributions. **Right;** Percentage of structure with a given number of interacting CDR.

2. Optimization and expression of the intrabody library

2.1 Objectives

In the previous works of our group, a human scFv, called scFv13R4, which is expressed at high levels in *E. coli* cytoplasm was obtained by molecular evolution (Martineau *et al.*, 1998). This scFv is also expressed under a soluble and active conformation in yeast (Visintin *et al.*, 1999) and mammalian cells (Sibler *et al.*, 2005). Additionally, scFv13R4 is very stable *in vitro* and can be renatured in presence of a reducing agent. Moreover, analysis of its folding kinetics showed that it folds faster than the parent scFv and aggregates more slowly *in vitro*. Based upon this scFv13R4, a human scFv library (PMEW) with more than a billion clones was constructed (Philibert *et al.*, 2007). Most of the scFvs in the library are expressed in *E. coli* and in mammalian cytoplasm, and are functional as intrabodies.

Founded on the results obtained with PMEW, a new human scFv intrabody (HUSCI) library was designed in order to improve the stability, diversity and affinity (Gautier *et al.*, 2014, in revision). This new library was also based on the hyper-stable scaffold scFv13R4 but strategic changes were made in its design. First, the diversity of the PMEW library is located exclusively in the two CDR3s in which all positions were mutated. In the new library, the variable positions are distributed throughout the 6 CDRs but correspond only to key positions in terms of antigen binding and protein stability (Figure 22). Secondly, the incorporation of amino acids in the variable positions of the PMEW library was modeled on the human natural distribution at each position. The use of 20 amino acids as mutation residues for the selected positions resulted in a diversity of about 10^{25} for the combinatorial library, which exceeds the maximum diversity of a phage library ($\sim 10^{10}$). In HUSCI library, one very restricted distribution is used for all positions. It includes the top 5 of the most frequently used amino acids at the intrabody-antigen interaction sites and can be subdivided in terms of ability to form specific connections (hot Y, D, N, and neutral G, S). Thirdly, in PMEW, the lengths of CDR3s were varied in accordance with the length distribution in human antibodies. In HUSCI, only the H3 length is variable and restricted to the 8 most common lengths that were incorporated with the same frequency (Figure 23). These differences in library construction allowed us to minimize the structural and stability discrepancies between HUSCI and scFv13R4.

Here at the Cancer Institute of Montpellier, several research groups made already use of the HUSCI library. Binders against their respective targets (p38 MAP kinase, Cathepsin-D, HER4, TAM family ...)

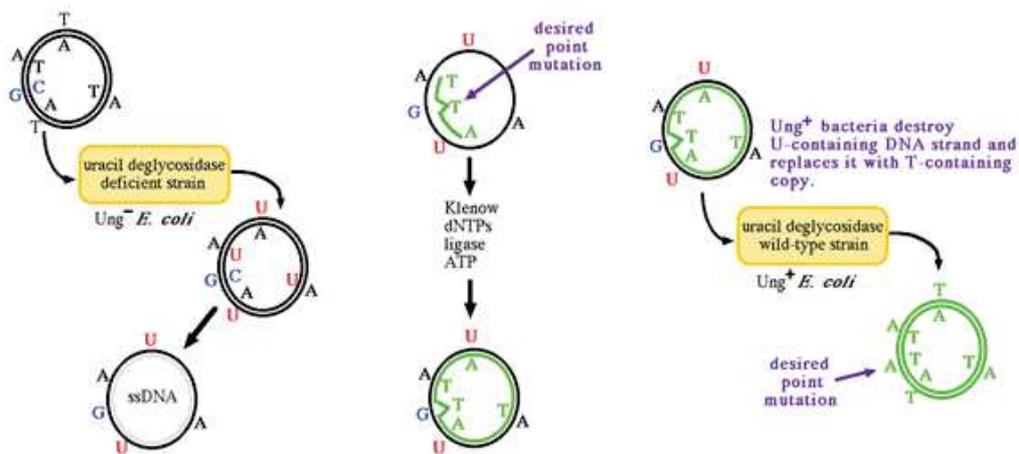


Figure 24: Kunkel mutagenesis (From <http://www.bio.davidson.edu/molecular/kunkel/kunkel.html>)

were successfully selected by phage display. Our project, however, aimed to use this scFv library in a totally different approach. Therefore, the library was redesigned (named HUSCIV) to be expressed via a retroviral system, for phenotypic screening in eukaryotic cells. In particular, my PhD project aimed to select for scFvs which restored the sensitivity to Cetuximab of *KRAS* mutated colorectal cancer cells. The intrabodies of interest are identified by two methods: by direct selection of cells blocked in their proliferation (by means of a fluorescent marker) and by high throughput sequencing of subsequent intrabody pools followed by a statistical analysis. To fulfill the constraints of this in cell screen, we had to adapt the construction of the library to match the experimental requirements.

2.2 Design of HUSCIV

The method for site-directed mutagenesis applied here, termed 'Kunkel mutagenesis' (Figure 24), features hybridization of a mutation-encoding oligonucleotide to a target site on a uracil-doped template plasmid (Kunkel, 1985). The oligonucleotide consists of a variable region sandwiched between two annealing sequences that are complementary to the target site. Phosphorylated oligonucleotides are annealed to the template plasmid, followed by *in vitro* enzymatic synthesis of the complementary DNA strand. The main advantage of the method is that the number of oligonucleotide is not limited and thus that several sites can be mutated in a single experiment.

The method can be divided in three steps. First, a uracil-doped single-stranded matrix is prepared using a *dut ung E. coli* strain. Second, oligonucleotides are hybridized to the matrix, extended with a polymerase devoid of strand displacement activity (T4 polymerase in our case). Of course, the *in vivo* synthesized strand contains T and not U bases. Third, the double-strand hybrid plasmids containing a wild type uracilated strand and a mutant non-uracilated strand are transformed in a wild type *dut+ ung+* strain. In this strain, the uracil-containing strand is destroyed and most of the clones contain a plasmid originated from the *in vitro* synthesized and mutagenized strain. Efficacy of the mutagenesis is usually about 70-80% per oligonucleotide.

The HUSCIV library is aimed to be used in eukaryotic cells and will be thus limited to a diversity of 10^6 – 10^8 depending on the project requirements. In the case of my PhD project and the selection scheme used for the reversal of Cetuximab sensitivity, we chose to limit the diversity to about 10^6 . It was thus of premium importance to optimize all the important steps in the Kunkel mutagenesis to obtain a small yet functional library.

This excluded the direct recloning of the previously constructed phage-displayed library (HUSCI) for two main reasons:

First, the phage-library was constructed to force the mutation of the three CDR loops. To achieve this goal, the mutagenesis was done on a CDR-free template, which represents scFv13R4 sequence lacking the CDRs, each of which being replaced by a single nucleotide. Replacing each CDR by a single nucleotide shifts the reading frame, thus introducing a STOP codon downstream of a non-mutated CDR and avoiding the display of scFvs containing any scFv13R4 CDR. Consequently, only 10-25% (1-0.7⁶; 1-0.8⁶) of the final library contains expressed clones since the efficacy of Kunkel procedure is about 70-80% per oligonucleotide. This is not a problem in the case of a large phage library (about 10¹⁰), but it is a limitation in the case of a very limited library like HUSCiv.

Second, the project seeks to identify intrabodies based on their declining frequency measured by next generation sequencing. To statistically and reliably identify declining sequences it is necessary to start with at least 100 copies of each intrabody. Since we seek to use a diversity of 10⁶, we need to sequence at least 10⁸ clones for each condition. Such a sequencing depth is only possible on an Illumina HiSeq2500 which is limited to a sequencing length of about 100 base pairs. The sequencing of a variable part of 550 bases (6 CDR domains) would have been possible but restricted to a diversity of about only 20 million sequences, too low for our experiment.

Therefore, a non-coding barcode of 30 random bases has been added to the library so they can be distinguished and sorted during data analysis (Figure 25). Because of the very large potential diversity of the barcode sequence (4³⁰ > 10¹⁸), much larger than the library diversity, a given barcode can only be associated with a single sequence of scFv. However, the reverse is not true and a given scFv can be represented by several barcodes. To monitor intrabody diversity in cell populations during the course of selection, their barcodes will be sequenced. Since the identified barcode are unambiguously associated with a single scFv, this will lead to the identification of the intrabody itself.

It must be noted that identification of a barcode will only theoretically lead to the intrabody sequence. In practice, we will have to rescue the full intrabody from the library in order to identify it. I had already shown that this can be successfully performed during my previous work on the phenotypic screening approach in the mast cells (Results: 1.4 Characterization of inhibitory intrabody families).

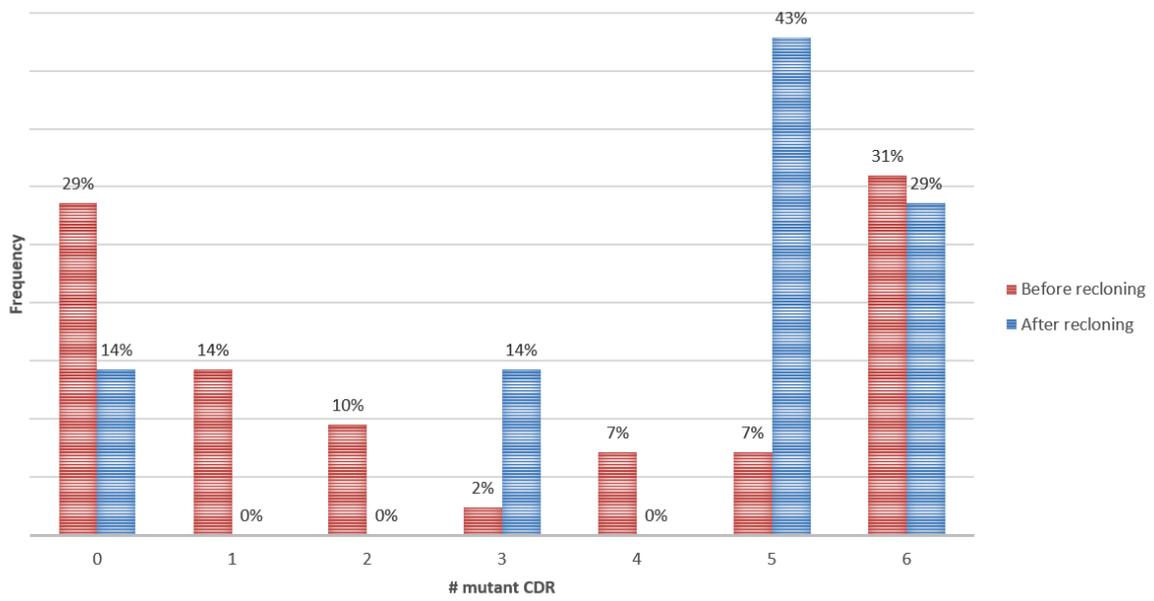


Figure 26: Average frequency for each number of mutant CDRs for a random scFv out of HUSC1v, before and after the recloning step into retroviral vector.

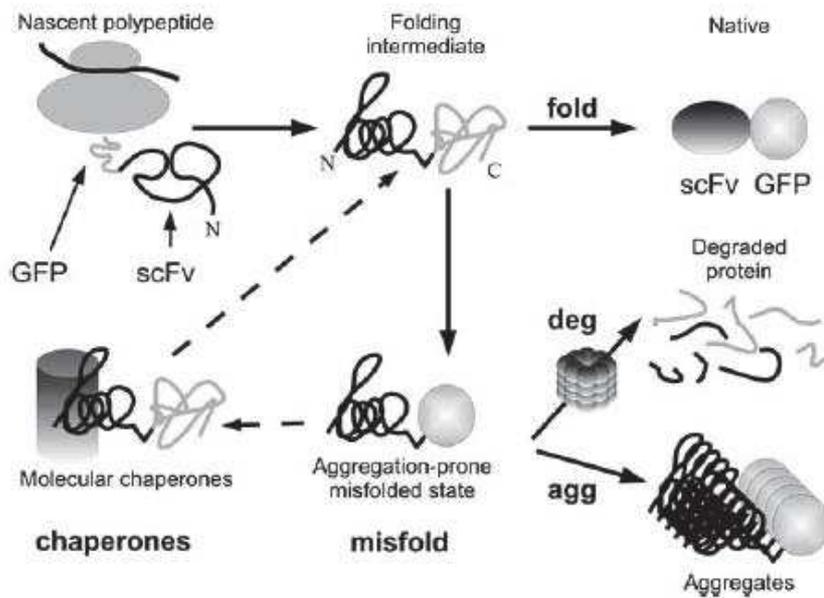


Figure 27: Schematic view of the *in vivo* folding fate of scFv-GFP fusions (Guglielmi et. al, 2011).

Indeed, I showed that starting from a barcode sequence (in that case the VH CDR3 sequence), it was indeed possible to design an oligonucleotide, amplify the VH and VL sequences and assemble them in a full scFv. The reconstituted gene could then be recloned in an expression vector. This allowed us to sequence the full scFv and to test its activity as an intrabody (Figure 17). Here, the same approach will be followed to identify and test the clones from the Cetuximab selection.

To address these two pivotal points in the construction of the HUSCiv library, we introduced the following modifications in the construction procedure:

We used a wild type scFv13R4 template instead of the CDR-deleted version. This implies that most of the clones will contain some non-mutated CDRs. However, the bioinformatics analysis of intrabody-antigen complexes developed for the construction of the original HUSC library showed that many antibodies use less than 6 CDRs to interact with the antigen (Figure 23). This restriction of the intrabody paratope diversity should thus not affect the functional diversity of the library. However, the first attempts to perform such a mutagenesis resulted in a strong bias because the oligonucleotides with a low mismatch frequency presented a higher affinity for the template than the more mismatched oligonucleotides. We thus modified the hybridization conditions of the annealing step. The idea was, once the denaturation step was finished, to lower the temperature as quickly as possible towards 20°C (10 seconds, temperature was set to 4°C) to avoid that the most complementary primers (high T_m) hybridize preferentially to the template because this would bias the library. Final sequencing of randomly picked clones showed that the procedure was successful and that no stop codons were observed (Annex 3).

Finally the library had to be cloned in a retroviral vector. We could have performed the mutagenesis directly in such a vector, however we decided not to do so to ensure that all the clones contain a barcode. Indeed, the barcode is introduced by the Kunkel procedure and thus 20-30% of the clones do not contain the barcode sequence. We thus constructed the library in a first vector then used a unique *BglIII* restriction site introduced by the barcode primer to reclone the library in the pMSCV retroviral vector. This approach not only ensured the presence of the barcode but also removed the non-mutagenized template originating from the non-eliminated uracil-containing template (about 30%). The total diversity of the HUSCiv library was determined to be at least 1×10^8 .

Before and after cloning, the variation in the CDR domains of both VH and VL was determined. Fifty clones were picked out for sequencing (Annex 3) and the frequencies of each CDR that has been mutated are shown in figure 26. One third of the scFvs out of the HUSCiv library are mutated in all six

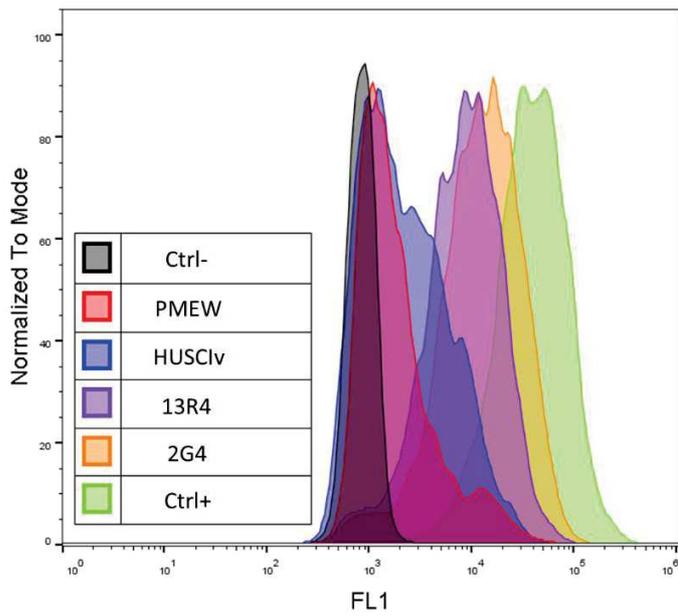


Figure 28: HeLa cells infected with retrovirus containing scFv library GFP fusions. Cell populations were analyzed by FACS using the GFP fluorescence signal.

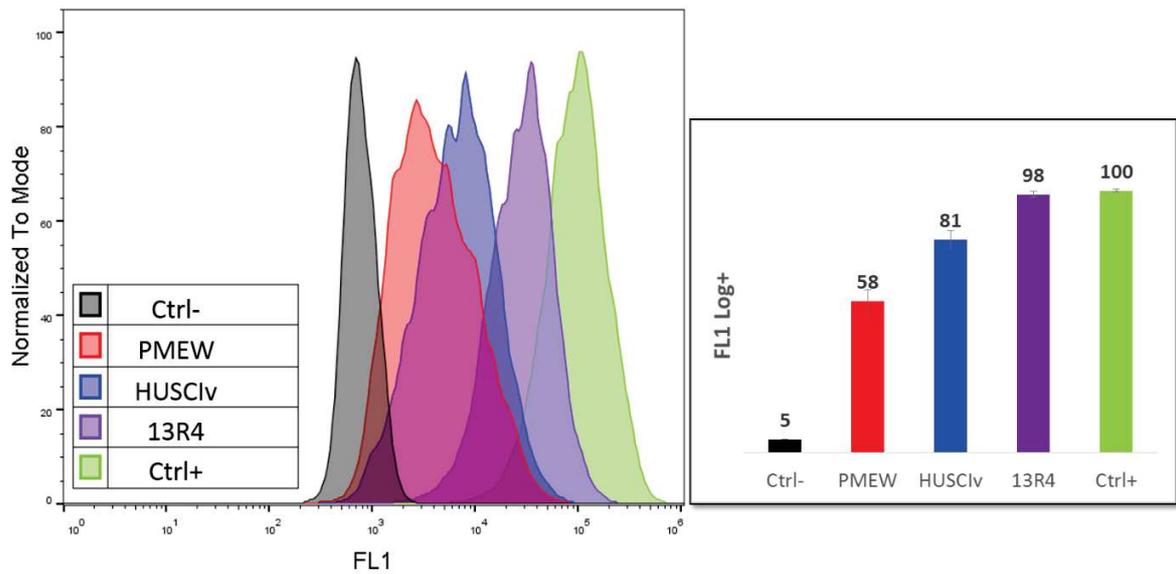


Figure 29: HCT116 cells infected with retrovirus containing scFv library GFP fusions. **Left panel:** Cell populations were analyzed by FACS using the GFP fluorescence signal. **Right panel:** FL1 log+ values measured as >5% GFP+ population of the negative control.

CDRs, and therefore represent a fully mutated template. Seventy percent of all scFvs in the library are mutated in at least one CDR domain. The number of scFv with at least 3 mutated CDR loops is 86 % in the final library. Since 97% of the natural antibody use between 3 and 6 CDRs (Figure 23) to interact with their cognate antigen, we can assume that most of the retroviral library is functional.

2.3 HUSCIV expression in eukaryotic cells

To evaluate their properties as intrabodies in mammalian cells both former (PMEW) and current (HUSCIV) retroviral scFv libraries were expressed in a frequently used human cell line (HeLa), and in our model cell line (HCT116). In an attempt to follow scFv stability and solubility in the mammalian cell cytoplasm, GFP from jellyfish was used as a folding reporter (Figure 27), as described by Waldo *et al.* in *E. coli* (Waldo *et al.*, 1999) and by our group in mammalian cells (Guglielmi *et al.*, 2011).

We first analyzed total GFP signal using FACS; the idea is that the GFP signal of the cell lines correlates with the soluble expression levels of the scFvs. For the controls, the lab previously described a scFv, called scFv2G4, selected for high expression levels in *E.coli* cytoplasm. It is an anti-tubulin scFv that once expressed in mammalian cells, accumulates as a soluble protein in the cytoplasm and of which no aggregated material was detected (Cassimeris *et al.*, 2013). The scFv13R4 which is the template scFv of both libraries was also included in our experiments. Additionally, an empty retroviral vector fused (control+) or not fused (control-) to EGFP were added to the experiments.

Figure 28 shows a clear shift (FL1 Log+ of 10) in the FACS signal of HeLa cells expressing PMEWE in comparison with HeLa cells expressing HUSCIV. The GFP levels were even more convincing in HCT116 cells: Figure 29 shows a clear augmentation of 23% of the GFP signal of HUSCIV fusion proteins in comparison with scFvs out of PMEWE.

However, the GFP signal can be due either to fusions or to cleaved and free GFP. We thus used Western blot of total and soluble cell extracts to analyze the expression levels of scFv fusions. Analysis of total extracts of HeLa cells showed that scFvs 13R4 and 2G4 fusions were both present at high levels in the cell cytoplasm (Figure 30, top panel; MW 60 kDa). On the other hand, due to the absence of a fusion protein, the empty control vector is expressed at much higher levels showing a high proportion of the EGFP protein (MW 28 kDa). But more importantly, cells expressing PMEWE also contains free EGFP protein, which is originating from degraded fusion protein whereas those expressing the HUSCIV library do not. This means that part of the signal associated with the PMEWE library in the previous FACS experiment (Figures 28 and 29) is not due to the intrabody itself but to

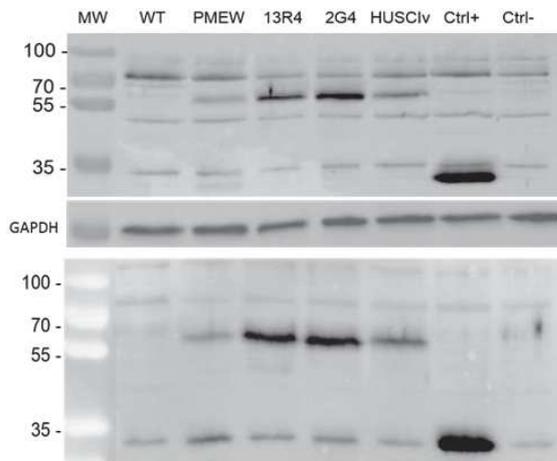


Figure 30: Western blotting of; **Top panel:** Total HeLa cell extracts revealed using a rabbit polyclonal anti-GFP serum; **Bottom panel:** scFv-GFP fusions were captured from soluble HeLa cell extracts by their C-terminal poly-His tag. WB was revealed using a rabbit polyclonal anti-GFP serum (1/500) coupled to an anti-rabbit HRP (1/5000). Exposure: 30 seconds.

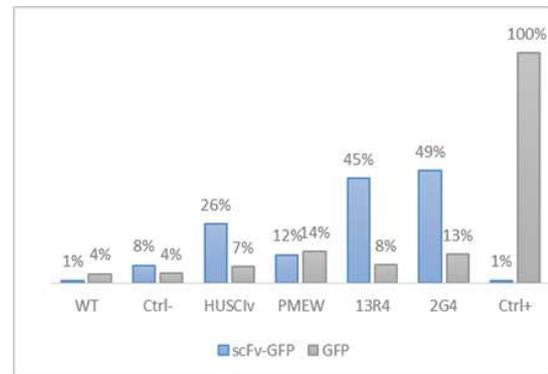


Figure 31: Western blot quantification of soluble cellular extracts of HeLa cells. Signal intensities are normalized to the positive control.

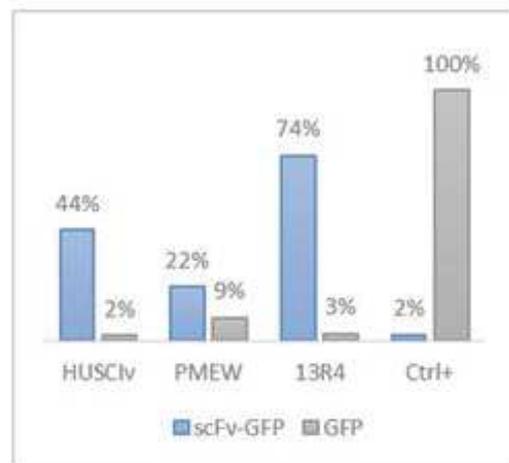
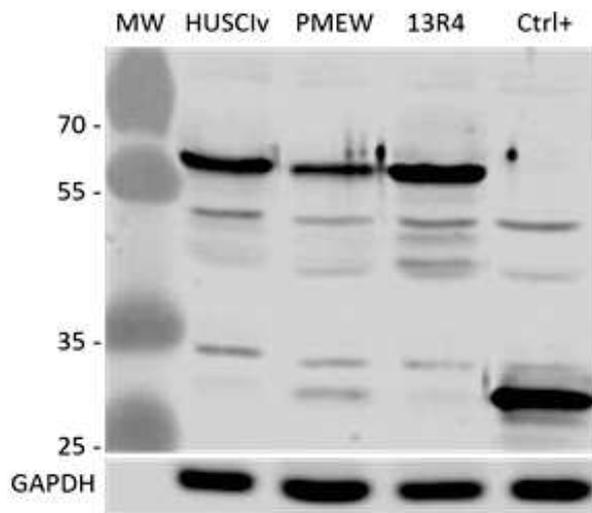


Figure 32: Western blotting of soluble cell extracts of HCT116 cells. **Left panel:** scFv-GFP fusions were captured from soluble cell extracts by their C-terminal poly-His tag. WB was revealed using a rabbit polyclonal anti-GFP serum (1/500) coupled to an anti-rabbit HRP (1/5000). Exposure: 30 seconds. **Right panel:** Western blot quantification of soluble cellular extracts of HCT116 cells. Signal intensities are normalized to the positive control.

cleaved-off eGFP. It also demonstrated that, in eukaryotic cells, the expression of the HUSCIV library is higher than the PMEWS library. However, to rule out the possibility of measuring scFv-GFP aggregates, soluble fractions of cell extracts were prepared and captured on Ni-NTA beads (Figure 30, bottom panel). The quantification of this western blot is shown in figure 31. In the case of scFv 13R4 and 2G4, both highly soluble fusion proteins are residing at high levels in the cytoplasm. The background was annotated through the GFP levels of our negative control (8%). Anyhow, we confirm that the proportion of degraded fusion protein in the PMEWS-transfected cells is 2 to 3-fold higher than in the HUSCIV-transfected cells. And more important, the amount of soluble scFvs in HUSCIV cells is 2 to 4-fold higher than in PMEWS library.

The Western blot analysis of the libraries expressed in HCT116 showed the same results (Figure 32). The template scFv13R4 fused with the EGFP is expressed at high levels (74%) in the cell cytoplasm. As we observed with HeLa cell extracts, the fraction of soluble fusion protein in cells expressing HUSCIV is twice as high as the fraction found in cells expressing PMEWS and represents 60% of the scFv13R4 control. In addition, we clearly notice the GFP degradation product in PMEWS-transfected HCT116 (9% versus 2% in HUSCIV).

Finally, to confirm the results obtained by FACS and Western blot analysis, we used fluorescence microscopy, to visualize the expression of the intrabodies. Figure 33 compares the overall fluorescence of HCT116 cells infected with PMEWS and HUSCIV. There is a much higher and uniform GFP intensity in cells expressing the HUSCIV library. Since we know that there is no degradation of the fusion (Figure 30 and 32), this represents expression of the intrabody itself as soluble protein in the cytoplasm. It is also important to note that no aggregates were present in the cells as frequently found with inactive and badly folded intrabodies.

Altogether, our data clearly demonstrate that most of the clones contained in the HUSCIV library are well expressed, non-degraded, and soluble in the cell cytoplasm. In addition the library is clearly better than the previously designed PMEWS library. This is true not only in HeLa cells but also in the HCT116 cell line that will be used for the phenotypic selection. This also demonstrated that the optimization steps developed in the construction of this new library were successful and HUSCIV represents a very promising tool for the generation of efficient intrabodies.

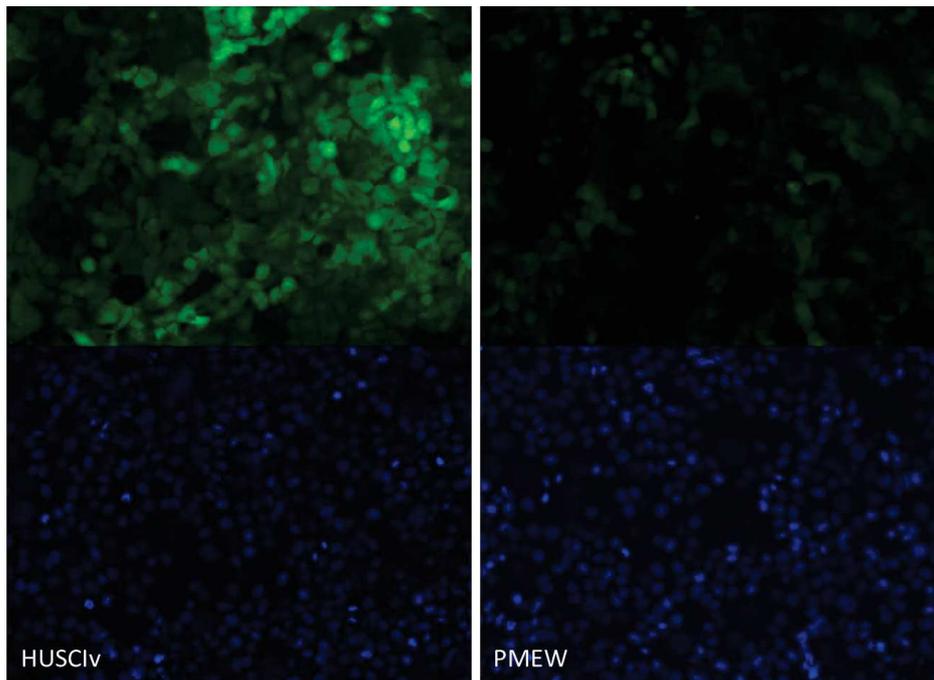


Figure 33: Expression of scFv in mammalian cytoplasm: HCT116 cells were retro-transfected with scFv GFP fusions and fixed by PFA. **Upper panel:** Intrinsic GFP fluorescence revealed by fluorescence microscopy (Plan-Neofluar 20x/0.50 – 1.18 s). **Lower panel:** Cell nuclei of respective cell area stained with Hoechst (Plan-Neofluar 20x/0.50 – 11 ms).

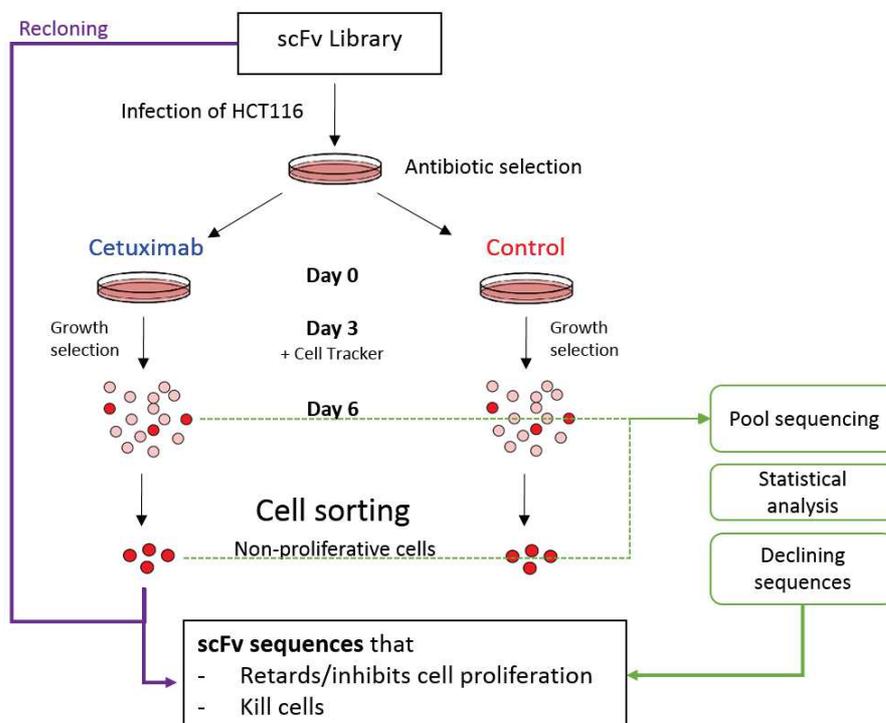


Figure 34: Phenotypic screen of for candidate intrabodies restoring the sensitivity of KRAS mutated colorectal cancer cells for Cetuximab treatment by means of retroviral transfection. Purple: Method 1, Green: Method 2.

3. Phenotypic targeting of KRAS mutated colorectal cancer cells

3.1 Objectives

The aim of the screen is to identify new therapeutic targets in *KRAS* mutated colorectal cancer cells by intracellular immunization. In analogy with the RNAi, the technique is known to express intrabody fragments in the cell to inhibit or modulate the functions at the protein level, and can therefore be called "protein interference". Usually, this technique starts with an antibody fragment against a chosen target protein. This intrabody is expressed by the cell in the form of a scFv, which then can modulate the function of the target protein by different mechanisms (direct inhibition, relocation, degradation, etc.). In the approach used here, it is the other way around: starting from the optimized HUSCIV retroviral scFv library, a screen for a desired phenotype will be performed and subsequently the intrabodies conferring this phenotype to the cells will be isolated. For my PhD project, I employed this strategy to select for intrabodies that restored the effect of Cetuximab on *KRAS* mutated colorectal cancer cells. Those intrabodies are identified by two methods (Figure 34):

- 4) Direct selection of cells blocked in their proliferation by means of a fluorescent marker.
- 5) High throughput sequencing of subsequent intrabody pools followed by a statistical analysis.

Once the intrabodies of interest are isolated, their potential therapeutic targets can be identified by mass spectrometry using intrabodies as conventional antibodies that immunoprecipitate the target as already demonstrated with the identification of C12orf4 in mast cells (see article Mazuc *et al.*, 2014, Results: 1.5 Target identification). Finally, the mechanisms of action of intrabodies will be characterized in the activation of the EGF-pathway in other cell lines treated or not treated with Cetuximab.

3.2 Experimental set-up

It should be noted that the newly synthesized retroviral intrabody library has a built-in diversity of 10^8 clones. The human genome contains about 21,000 protein-encoding genes, but the total number of proteins in human cells is estimated to be between 250,000 to one million. We thus used a subset of the library of approximately 10^6 clones. This diversity is of the same order of magnitude as the proteome diversity taking into account alternative splicing and post-translational changes. In addition, as explained (2.2), this diversity is the largest possible size to be followed by NGS.

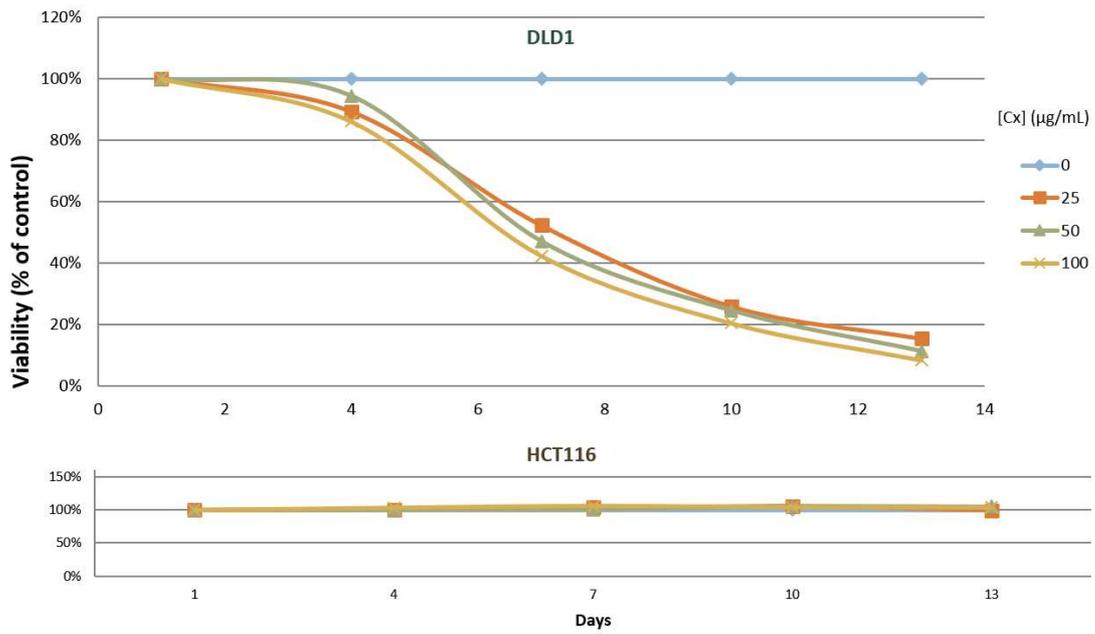


Figure 35: Cytotoxicity of Cetuximab-treated colorectal cancer cells.

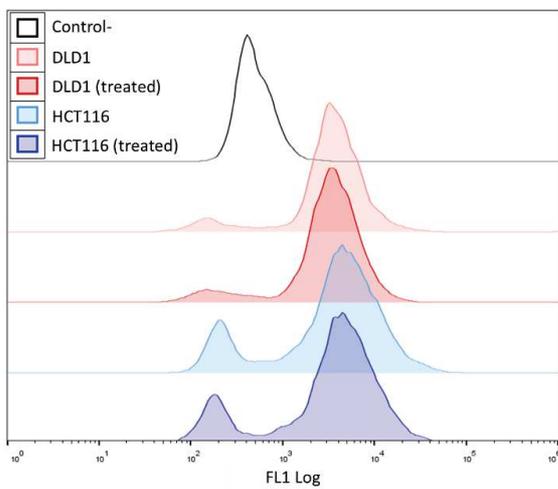


Figure 36: EGFR internalization of Cetuximab-treated colorectal cancer cells. Fluorescence of treated and non-treated cells was measured by FACS after incubation with Cetuximab (50 µg/mL) followed by an Fc-specific anti-human IgG – FITC. Negative control: Non-treated HCT116.

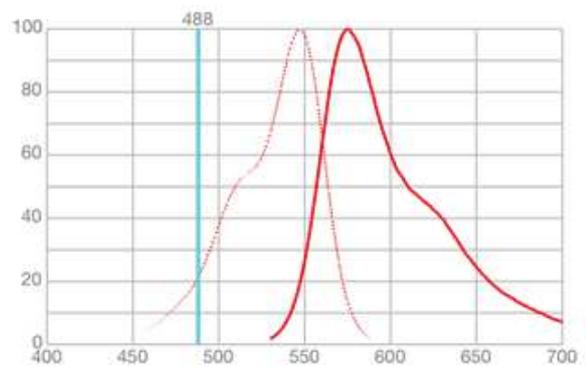


Figure 37: Dichromatic emission spectra of green EGFP (dotted line) and Cell Tracker Orange (full line)

First, we tested the cytotoxic effect of Cetuximab on CRC cell phenotypes. Oncogenic activation of the RAS/RAF signaling pathway significantly impairs the therapeutic potential of mAbs such as Cetuximab aimed at targeting the EGFR in CRCs. Two CRC cell phenotypes, carrying the same *KRAS* mutation Gly¹³Asp *KRAS* (Benvenuti *et al.*, 2007). For these experiments, we selected the DLD1 cell line, which is known to be partially Cetuximab-sensitive and HCT116, a resistant cell line. The cells were treated with increasing concentrations of Cetuximab and their viability was measured every 3 days during two weeks (Figure 35). As expected, HCT116 cells do not respond to Cetuximab treatment but after 1 week, the cell viability of DLD1 cells was reduced to 50% of the non-treated control. We observed that an incubation time of at least 6 days is necessary to get a strong cytotoxic effect on Cetuximab-sensitive cells. Moreover, the cytotoxicity of DLD1 measured after 6 days of Cetuximab treatment is nearly the same for all Cetuximab concentrations.

In order to exclude the possibility that HCT116 cells internalize their EGFR after being treated with Cetuximab for 6 days, a FACS analysis was performed. HCT116 and DLD1 cells were incubated with or without 50 µg/mL Cetuximab for 6 days and EGFR levels were measured by FACS by means of an Fc-specific anti-human IgG. Despite a slightly larger number of unmarked cells in comparison with DLD1, a same level of fluorescence was assessed for both treated and non-treated HCT116 (Figure 36).

Eventually, we had to test the concentration of the fluorescent dye used for the proliferation test. The CMRA Cell Tracker™ fluorescent probe freely pass through cell membranes and is converted to a cell-impermeant product. The dye resides in the cytoplasm and is passed to daughter cells through several generations but is not transferred to adjacent cells in the population. Cell blocked or altered in their proliferation will be more fluorescent than adjacent cells that maintain their normal growth. As such, those cells can be positively sorted by FACS. However, as successfully transfected cells are expressing a scFv-EGFP fusion protein the FACS instrument has to be tested for dichromatic analysis with green (EGFP, FL1) and orange (CMRA, FL4) overlapping emission spectra (Figure 37).

Therefore, a FACS analysis of was performed with increasing concentrations of fluorescent dye. Cells were transfected with an irrelevant scFv-GFP and incubated for three days with CMRA concentrations varying from 0 to 20 µM. FACS analysis annotates cells bearing a GFP-labeled intrabody in their cytoplasm due to successful transfection and the cellular proliferation state represented by the final concentration of CMRA. The dichromatic issue resides in the orange component, which is displaying a “green component” - shifting the cell population to the right (more GFP-positive) - too intense to be removed by color compensation. The aim was to decrease the CMRA concentration until this component was low enough to be corrected and one would analyze both fluorescence

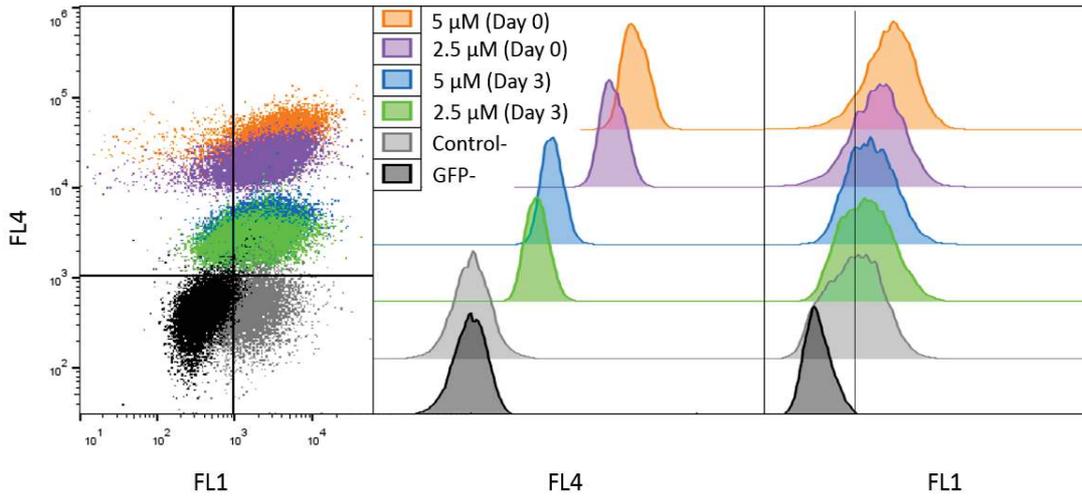


Figure 38: FACS analysis of scFv-GFP transfected HCT116 incubated with increasing concentrations of CMRA. **Left panel:** Total cell populations. **Center panel:** CMRA component. **Right panel:** GFP component. Cells were incubated with 0 to 20 μM of CMRA (for clarity, by comfort, only the two best concentrations of 2.5 μM and 5 μM of CMRA data are shown). The negative control (Control-) is represented by unstained scFv-GFP transfected HCT116. Unstained Wild Type cells served as GFP-negative control (GFP-).

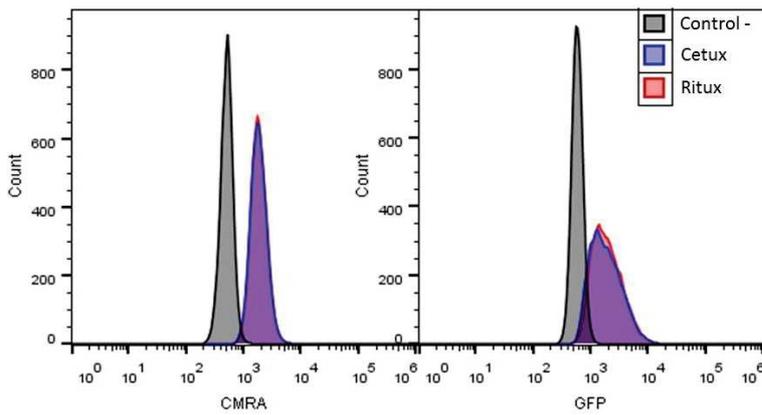


Figure 39: Effect of monoclonal Abs on FACS analysis of scFv expressing HCT116. **Left panel:** CMRA component. **Right panel:** GFP component. Cells were incubated with 0 to 20 μM of CMRA. Antibody concentrations were set at 50 $\mu\text{g}/\text{mL}$. The negative control is represented by unstained wild type cells.

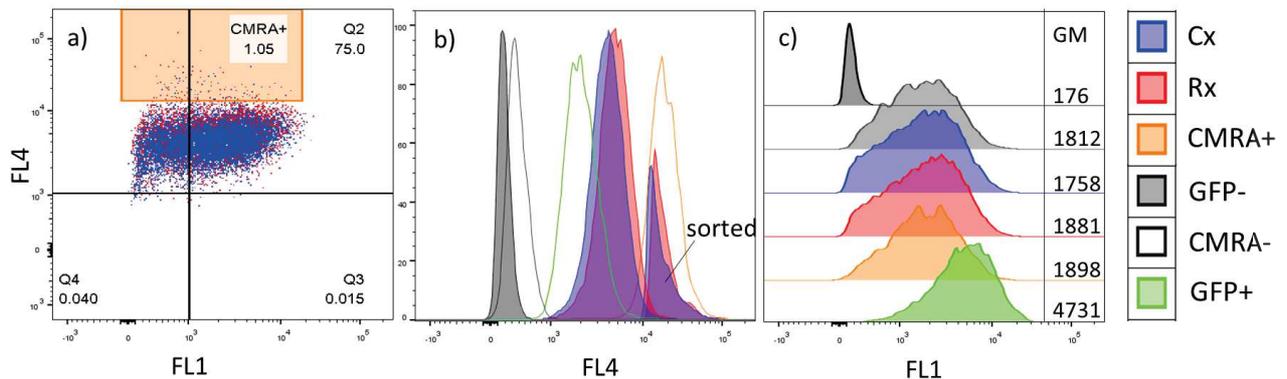


Figure 40: FACS analysis of HUSCIV transfected HCT116 **a)** Sorting window of cells retarded in their proliferation, treated with Cetuximab (blue) or control (red) for 6 days. **b)** FL4 component (CMRA): Sorted subpopulations are depicted in their respective colors. **c)** FL1 component (GFP) with their respective Geometric Mean (GM). Controls: **CMRA+**: HUSCIV transfected cells at maximum staining; **GFP-**: Wild Type cells; **CMRA-**: Unstained HUSCIV transfected cells, **GFP+**: Cells transfected with an empty GFP-expressing retroviral vector.

components in their true condition. However, the concentration of CMRA should still be able to distinguish the state of proliferation in the population; i.e. proliferation arrest (represented by “Day 0”) and normal proliferation (represented by “Day 3”) (Figure 38; left and center panel). In addition, we should be able to visualize the difference between stained and unstained (negative control) cells. In the end, we have set the CMRA concentration to 2.5 μM since the use of a 2-fold higher CMRA concentration alters the GFP levels significantly (Figure 38; right panel).

To ensure that the addition of monoclonal antibodies to CMRA-stained cells did not have any effect on cell viability, a FACS analysis was performed on stained cells transfected with an irrelevant scFv-GFP and treated with Cetuximab and Rituximab for 6 days. It is known that HCT116 cells do not express CD20, which is the target of Rituximab. No differences in cell proliferation were noticed (Figure 39).

Based on these results the final parameters of the intrabody-based phenotypic screen of HCT116 cells treated with Cetuximab were fixed to 50 $\mu\text{g}/\text{mL}$ of Cetuximab during 6 days using a CMRA concentration of 2.5 μM for the last 3 days. In the best case represented by DLD-1 cells, we could get a 50% decrease in cell proliferation.

3.3 Phenotypic screen

As depicted in figure 34, the goal of the experiment is to identify and isolate scFvs that inhibit or decrease the proliferation of HCT116, in presence of Cetuximab.

First, about 7 million HCT116 cells were transfected with a subset of the HUSCIV library containing 10^6 intrabodies. The multiplicity of infection (MOI) was set to 1:1 in order to minimize double transfections. The transfection was successful as measured by the geometric mean (GM) of the GFP fluorescent signal (Figure 40c). Secondly, cells were incubated for 6 days with 50 $\mu\text{g}/\text{mL}$ of Cetuximab. Three days before analysis, cells were stained with CMRA and re-cultured. At day 6, cells were sorted out by FACS. The sorting parameter for CMRA-positive cells was set to 1% of the total sorted population (Figure 40a). However, 1 week of anti-EGFR treatment did not show a measurable difference in cell proliferation, as analyzed using CMRA staining (Figure 40b).

Following this first selection round, we introduced a recloning step. This ensured that the retrovirus-induced phenotype was associated with the expressed intrabody sequence and was not due to a cell drift or a particular genomic insertion site. Half a million of sorted cells were brought back in culture, DNA was extracted, and the intrabodies (together with the GFP and the barcode) were recloned in the retroviral vector. A second round of selection was then performed under the same conditions as

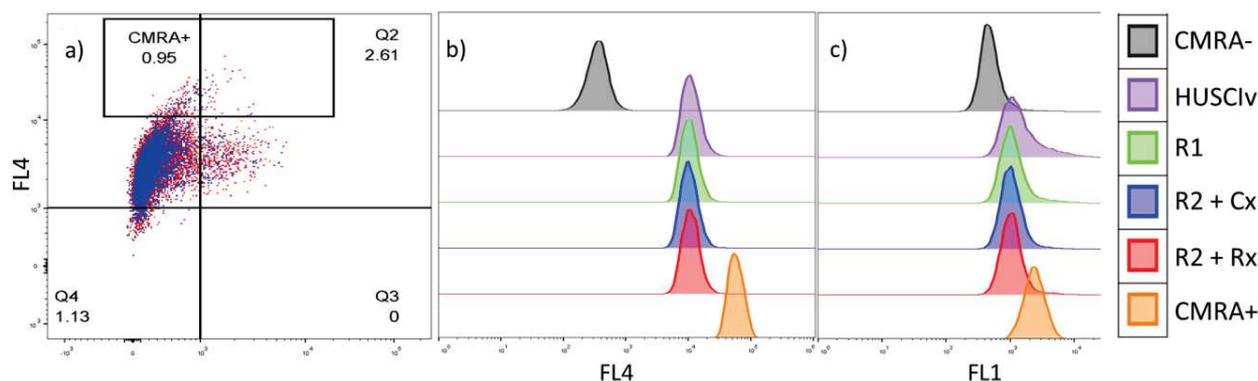


Figure 41: Round 2 of FACS analysis: HCT116 cells transfected with Round 1 were treated for 6 days with 50 $\mu\text{L}/\text{mL}$ Cetuximab and stained with CMRA Cell Tracker (FL4) 3 days before analysis. a) Sorting window of Round 2: about 1% of highest CMRA-positive cells. b) CMRA component (FL4): Synthetic representation of the proliferative state of Round 2 sorted cells compared to that of the naïve library (purple) or Round 1 (green). c) GFP component (FL1): idem. The negative control is represented by unstained cells (CMRA-), the positive control (CMRA+) by cells stained just before analysis.

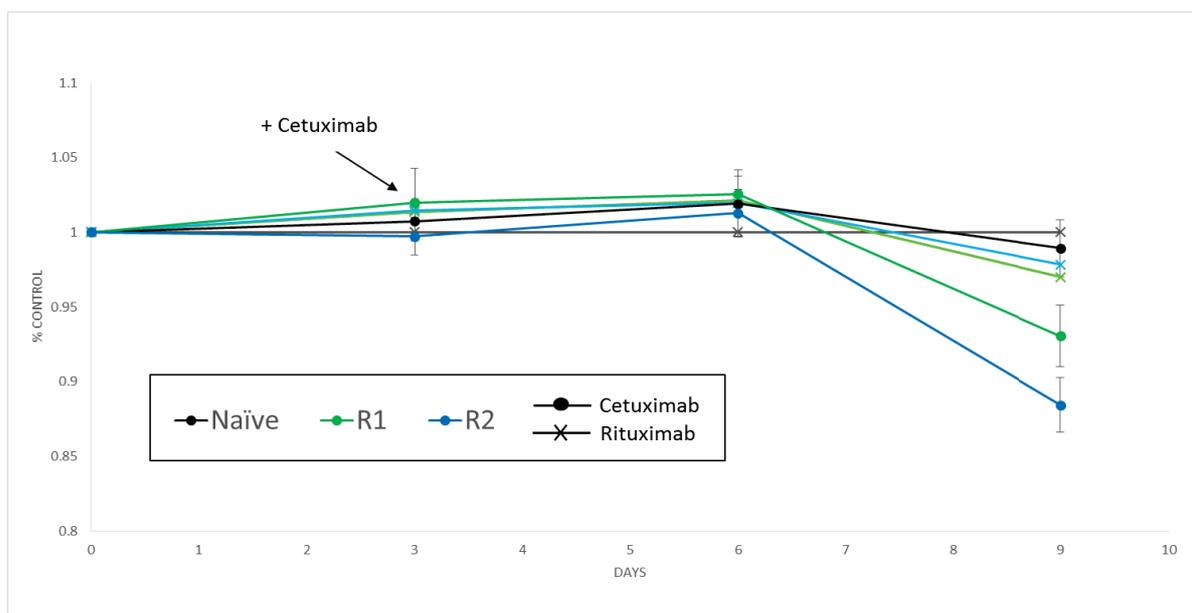


Figure 42: Cytotoxicity of HUSCIV-transfected HCT116 after two rounds of phenotypic selection with Cetuximab. Round 2 (R2) cells represent a subpopulation of Round 1, which have been sorted out by FACS and returned in culture. Round 1 cells were recloned after the first round of phenotypic selection. All cells were cultured for 3 days before adding 50 $\mu\text{g}/\text{mL}$ Cetuximab. Cell viability was analyzed every 3 days by a Sulforhodamine B assay. Values are normalized to cells transfected with the naïve library, treated with control mAb (Rituximab).

before except that the infection was performed at a MOI of 0.1 to further limit multiple infections. As expected, cell sorting has decreased the diversity of the subpopulation as illustrated by the wideness of the GFP+ distribution in the total cell population (FL1, Figure 41c). We estimate that the number of potential inhibitory scFvs is presumably between 10 and 10000, so maintaining 10^5 sorted cells should be enough to fish most of them. For this reason the second round was performed on 10 million of cells instead of 50 million. With a sorting parameter of 1%, we finally selected 100,000 clones upon 10 million cells (Figure 41a). The second round of phenotypic screen did not show any inhibition of the phenotype. After 3 days of CMRA staining, both “Round 1” and “Round 2” populations are superimposed on the naïve library, with or without 6 days of Cetuximab/control treatment (Figure 41b). To conclude, CMRA staining in combination with FACS analysis failed to show any difference in the proliferation status between HUSCIV-transfected cells treated with Cetuximab and control.

We surely can speculate that, due to false positives, 2 rounds of phenotypic selection are not sufficient to annotate a clear shift between cell populations. Another hypothesis is that the experiment is not refined enough to distinguish between CRC cells altered in their proliferation due to concomitant inhibiting effects of Cetuximab and scFv interaction, and the negative cellular effects of expressing intrabodies in the cytoplasm.

3.4 Cytotoxicity experiment

To verify that we did not miss out any inhibitory effect of Cetuximab on CRC cells during FACS analysis, we used an alternative cell proliferation assay. Aliquots of cells from the first and second selection round were seeded at limiting dilutions and a cytotoxic test was performed in quintuplicate by measuring the total cell mass (Sulforhodamine B assay, Figure 42). Cells were incubated for 3 days before adding 50 µg/mL of Cetuximab and control for 6 more days. Contrary to the FACS analysis, the proliferation of cells sorted out of the first selective round treated with Cetuximab was significantly inhibited by 5% on average when compared to cells treated with control mAb. More importantly, cells resulting from Round 2 showed a 10% difference in viability when treated with Cetuximab. The experiment was standardized towards the naïve library treated with Rituximab, which is a double control. First, by transfecting cells with the naïve library, we could interpret the effect of Cetuximab on - and its evolution within - a population of inhibitory scFvs. A second dimension (the effect of a monoclonal antibody) was added by incubating the cells with Rituximab, which does not to interact with HCT116 cells. The negative outcome of those 2 controls confirms the specificity of the measured effect and showed that the retroviral selection was more powerful than expected by FACS analysis.

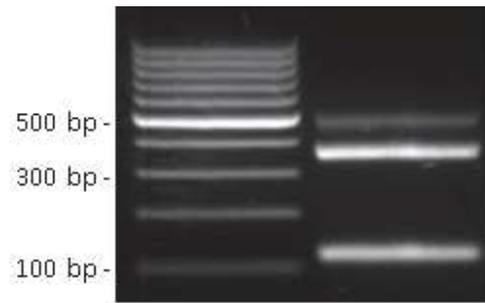
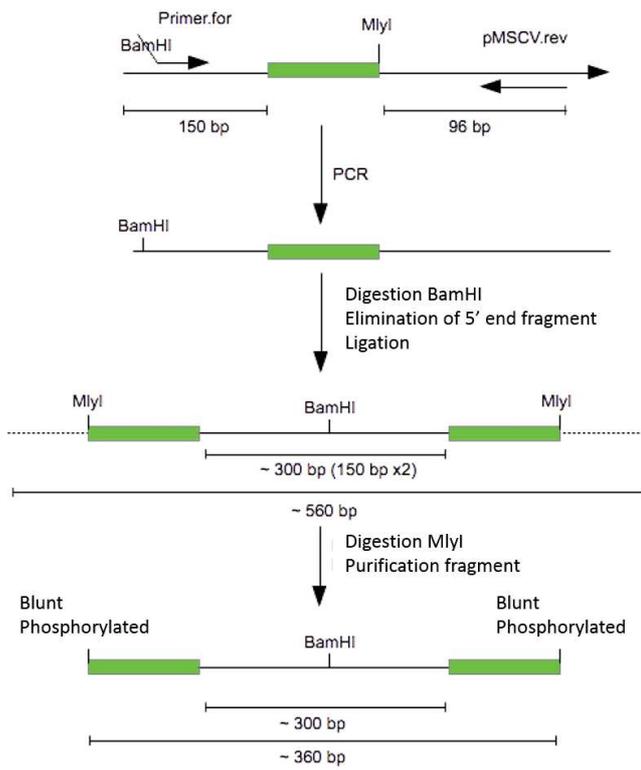


Figure 43: Preparation of fragments to be sequenced. **Left:** Schematic overview. **Right:** Agarose gel with dimer (560 bp), final sample (360 bp) and the MlyI digestion products (116 bp)

The differences in the two experiments may be due either to a lower sensitivity of the FACS/CMRA analysis or to the fact that the Sulforhodamine-based assay measure not only cell proliferation but also survival and cloning efficiency. These results are promising, especially because the effect increases with the selection rounds, and because of the strict negativity of all the performed controls. However, several additional rounds of selection will be necessary to obtain efficient intrabodies and confirm these preliminary results.

3.5 Sequence analysis

The experiment was performed in triplicate, which is necessary for statistical analysis. Intrabody sequences expressed in both pools of HUSCIV-infected cells treated with Cetuximab or Rituximab, before and after cell sorting were analyzed (see Figure 34). At “Day 6”, we anticipate the scFv diversity should be around 10^6 , with 0-1000 copies of each barcode. We need at least 10^8 sequences to have a precise description of the population, which is an average sequencing depth of 200. We thus sequence the pool of intrabodies before and after cell sorting under both treatments, and identify the declining sequences.

The barcode contains a *MlyI* restriction site, which enables the preparation of the fragment to be sequenced. First, the scFv was amplified out of the genomic DNA of HCT116. A second PCR step amplifies the barcode region and introduces a single BamHI site, which was used to create a dimer (Figure 43). After digestion with *MlyI*, the final fragment obtained contains a barcode at each end. *MlyI* is cutting blunt, so the fragment is phosphorylated and ready to be ligated with Illumina adaptors. Further development of the library is performed by using Illumina’s TruSeq[®]ChIP Sample Preparation kit and comprises 5 steps:

1. DNA end repair and phosphorylation. After testing both condition (with and without DNA end repair), we omitted this step because of several embedded PCR fragments of aberrant size.
2. Addition of adenine at the 3’ extremities.
3. Ligation of adaptors at both extremities.
4. Gel purification allowing to select the correct size (500 bp +/- 50).
5. An amplification step by PCR. Only fragments carrying both adaptors are amplified, introducing a second selection step.

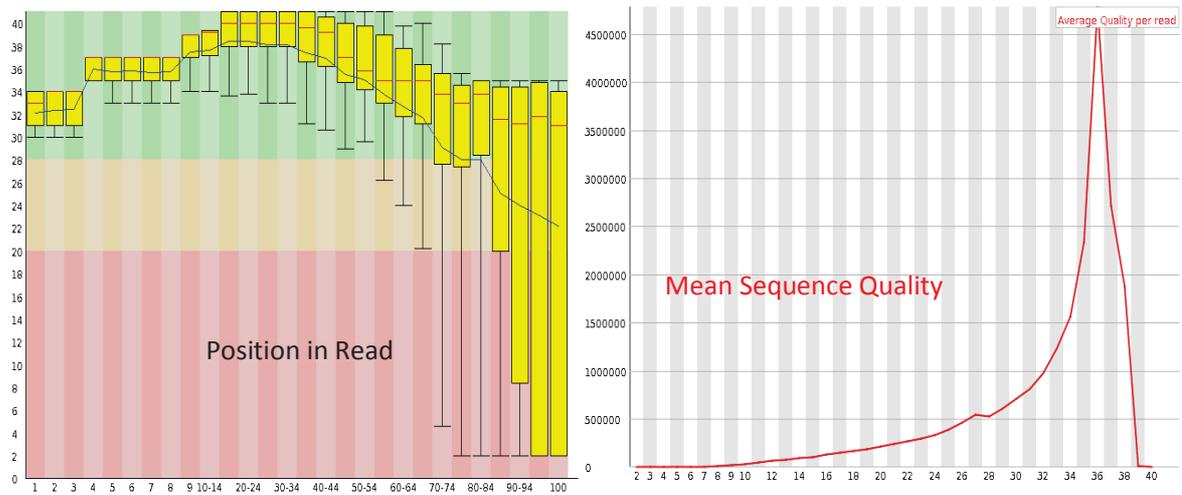


Figure 44: Quality of reads. **Left;** distribution of the quality scores for each base. **Right;** distribution of the mean quality for each sequence.

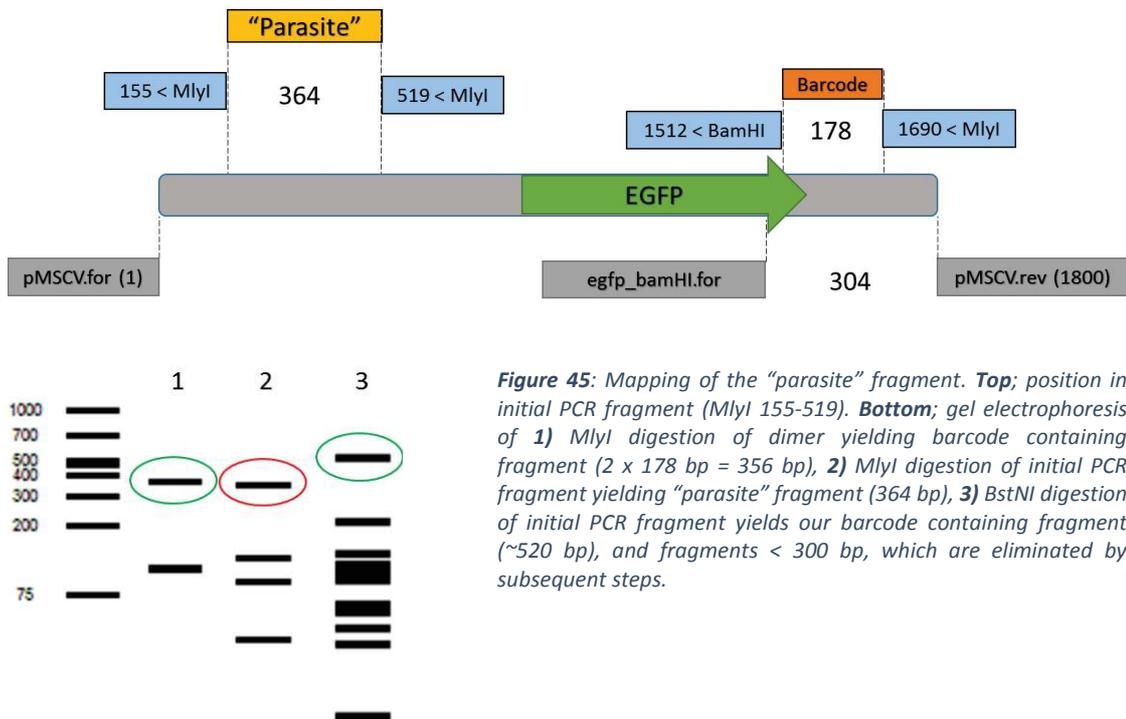


Figure 45: Mapping of the "parasite" fragment. **Top;** position in initial PCR fragment (MlyI 155-519). **Bottom;** gel electrophoresis of **1)** MlyI digestion of dimer yielding barcode containing fragment ($2 \times 178 \text{ bp} = 356 \text{ bp}$), **2)** MlyI digestion of initial PCR fragment yielding "parasite" fragment (364 bp), **3)** BstNI digestion of initial PCR fragment yields our barcode containing fragment ($\sim 520 \text{ bp}$), and fragments $< 300 \text{ bp}$, which are eliminated by subsequent steps.

The library construction and NGS experiments were performed on Montpellier MGX platform. The sequencing itself was performed on Illumina's *HiSeq 2000*, which makes use of the SBS technique (Sequence by Synthesis). This strategy is known as "bridge amplification" and generate "clusters", which are detected and DNA sequence determined via sequential incorporation of fluorescent nucleotides and a CCD camera. Before sequence analysis of all samples (3 x 2 conditions of input sample + 3 x 2 conditions of sorted sample), a single test sample was prepared and sequenced. The distribution of the quality scores of the sequences is depicted in Figure 44 (left). We observe a decline in quality at the end of the reads (especially the last 10 bases). However, only the 30 first bases of those reads are kept after trimming and a 10% error is allowed for the 20 fixed bases following the random barcode (i.e. authorizing 2 mismatches). Figure 44 (right) shows that the vast majority of the sequences is of excellent quality. After extraction of the barcodes, it is expected to have the following construction with on the first 30 bases, a barcode that represents only one scFv sequence, followed by 70 base pairs of a known sequence:

Nx30CAACAGTCTATGCGGCCCATTCAGATCCTCTTCTGAGATGAGTTTTTGTCTGCGGCCCGTGATGGTG

The expected "barcode" diversity is around 1 million. Finally, 5,466,475 reads contained the known barcode sequence, followed by the first 20 bp that corresponded to 24.8 % of the reads. 1,040,131 different barcodes were found, which corresponds to the expected diversity and 147286 barcodes frequencies are superior to 10. However, only 1/4 sequences are representing a barcode. After "BLASTing" the overrepresented sequences, we found out that a "parasite" sequence with the same size as our fragment, and originating from a *MlyI* digestion of our initial PCR fragment, was contaminating the sample. Normally this *MlyI* fragment should have been eliminated, but must have been carried through the purification as a non-specific PCR product. Therefore, an additional preparation step was introduced. For this, we digested the initial PCR product with *BstNI*. The enzyme cleaves many times, including 3 in the unwanted *MlyI* fragment and does not cut into the fragment containing the barcode (Figure 45). We assumed that by adding this digestion step, the contamination will be eliminated. Of course we risk to eliminate some barcodes (containing *BstNI*) but this should not bias too much the experiment hence among the 1 million of read barcodes, we expect to have only 54,719 (5%) containing a *BstNI* site. Recent results revealed an increase in the barcode reads to 42%.

Today, the major contamination problems has been solved and all samples are ready to be sequenced. Since the power of the HiSeq is 2×10^8 , if we get about 50% of barcodes, we can retrieve 1×10^8 sequences (100 times the diversity).

Conclusions

Molecules with the potential to recognize, bind and modulate intracellular targets play an important role for basic research in cell biology as well as in the generation of therapeutic and diagnostic agents. For real time visualization of cellular components, intracellular target identification and validation in living cells, such binding molecules basically have to fulfill three criteria: compactness, folding stability, and specific affinity. During the last ten years, elaborate research on recombinant antigen binding proteins, the development of combinatorial libraries, screening methodologies and cellular targeting technologies led to the identification and creation of numerous intracellular functional binding molecules (intrabodies). Selected intrabody formats were shown to function for different applications including target inhibition or modulation, studying protein–protein interactions or to visualize dynamic changes of cellular components.

In collaboration with E. Weiss' team, our lab constructed a intrabody library based on a single scFv framework, which was evolved to improve its activity inside the cytoplasm (Martineau *et al.*, 1998). The parental scFv13R4 presents favorable folding and aggregation kinetics and is expressed at very high levels in all tested cell types (Visintin *et al.*, 1999) (Philibert *et al.*, 2007). The diversity of this scFv library (PMEW) was restrained to the CDR3 loops, and was sufficient to generate very efficient and diverse binders (Philibert *et al.*, 2007) (Desplancq *et al.*, 2011) (Rinaldi *et al.*, 2013). Thus, by mimicking the natural diversity of CDR3 loops we ensured that the scFvs were fully human and functional in the cell. However, as it is known that all six CDRs play a role in intrabody specificity and affinity, we reconsidered the design of the intrabody library. The goal was to optimize the diversity, the expression and the stability of the library expressed in reducing intracellular conditions. We based our strategy on a database of crystallographic intrabody-antigen structures, in which an alanine scanning analysis taught us which characteristic rules had to be applied. We found that 80% of the binding energy is due to only a few significant residues. These residues localize among 25-30 fixed hot spot positions that contribute to 75-80% of the energy in more than half of the complexes. In addition, between 3 and 6 CDRs are involved in the complex formation. We thus restrained the diversity to the identified positions and focus us on the most widely used side-chains. In addition, we introduced variable loop lengths in the VH CDR3 to generate different paratope shapes. However, we restrained this diversity also to preserve as much as possible the initial stability and expression level of the 13R4 framework. Altogether, this allows the construction of smaller yet efficient intrabody libraries.

Successful use of scFv fragments in intrabody libraries on a large scale requires several primary points to be fulfilled. First, the scFv must be easy to isolate. Second, the scFv should be able to fold in all the cell compartments, particularly in the reducing ones. This is the case for the scFv library described herein, as represented by the quantity of soluble fraction residing in mammalian cells expressing the HUSCIV library. The comparison between PMEWS and HUSCIV demonstrates that the new library has improved expression and folding properties. Using GFP as a folding reporter (Guglielmi *et al.*, 2011), we clearly demonstrated that most of the intrabodies from HUSCIV library are expressed and soluble in the cell cytoplasm. Of particular interest is the homogeneity of the fluorescence (Figure 33) that ensures that the expression level of all the clones is comparable, thus limiting the risk of artifacts during the selection. Previous data showed that PMEWS contains about 30% of, which is a highly soluble scFv. The newly synthesized library contains two times less of scFv13R4, but is still more soluble than the former. This means that the introduced mutations in the 6 CDR loops did not reduce the overall solubility of HUSCIV.

In this PhD project, both intrabody libraries were applied in the phenotypic screening of well-documented phenotypes.

In the model of allergy, the PMEWS library was engaged in a mast cell activation model, which permitted us to identify a new molecular actor involved in this signaling pathway (these results have been recently published in PLoS ONE journal, see article Mazuc *et al.*, 2014). Such proteins of unknown function represent about 40% of the human genes. The identified LOC297607/C12orf4 protein is even part of the 1,000 human genes without ortholog, paralog, or homology to known genes (Clamp *et al.*, 2007). This class of uncharacterized proteins represents a large reservoir of therapeutic targets but because of their unknown properties they also represent a challenge for inhibitor identification (Mayr and Bojanic, 2009). This selection is a unique example in the current literature.

Of interest is also the recent publication by Dr. Lerner's group of a comparable selection. In this publication, they selected for intrabodies that block viral replication and thus protect cells from killing. This clear-cut phenotype (survival/killing) allowed them to get very high enrichment factors (Xie *et al.*, 2014). However, our selection is based on a much more general phenotype that could be applied to many more systems in which the phenotypes are usually less pronounced.

For my main PhD project, I used the new optimized library (HUSCIV) to select for intrabodies that restore the sensibility of *KRAS* mutated colorectal cancer cells to Cetuximab treatment. The selection

process for both projects is based on high throughput sequencing of scFv populations evolving throughout the experiment. Even so, one crucial dissimilarity has to be pinpointed. In the case of the screen performed in mast cells, the desired phenotype (inhibition of degranulation) keeps cells in culture (enrichment) whereas in the CRC model, the number of successful intrabodies (inhibited cell growth or death) will decrease or even completely disappear during the selection (depletion). Thus the latter requires a profound sequencing of a large number of scFvs, sufficient for statistical analysis. Additionally, because the variable parts are distributed along all CDR loops, the fragment makes up to 550 bases, obliging us to sequence through barcodes. Because we wanted to identify cytotoxic intrabodies, a total of 10^8 intrabody sequences (with about 100 copies/intrabody) were compared among two groups (before and after treatment with Cetuximab) and sequences specifically depleted during the selection could be discovered. Currently, the different DNA pools are ready to be sequenced and the sequencing of our test sample revealed the expected barcode diversity of 10^6 . Unfortunately, our first sequences showed a contamination, and our samples required additional purification. These steps have been applied and recent sequencing revealed that 43% of our samples contained a correct barcode. Notwithstanding the hold back, the phenotypic screen described here in the CRC model is known to be unique and promising.

In addition to the sequencing strategy applied on our CRC model, we performed a direct selection of the cells using a fluorescent dye (CMRA). Phenotypic screening by use of CMRA as a fluorescent arbitrator is well-documented and can track each cell division with a good resolution. In Cetuximab-sensitive cells used as control, we observed a difference of about 50 % in proliferation over 6 days of Cetuximab treatment. However, after two rounds of selection in the presence of Cetuximab, we did not observe any differences in proliferation rates, probably too weak at this stage to be annotated by FACS. In regard to this hypothesis, one should note that compared to the high throughput sequencing, the fluorescent marker approach does not identify cytotoxic intrabodies but only the cytostatic ones. Our results suggest that we need to add at least one or more selective rounds before formulating any conclusions on the approach. In that regard, a similar technique was successful in which cytostatic aptamers were isolated using CMTMR (de Chasse *et al.*, 2007). In this study it was stated that because of a significant background of cells that either do not grow or proliferate more slowly independently of the expression of peptide aptamers, multiple screening rounds were necessary to isolate peptide aptamers that exert an anti-proliferative effect. In the end, 7 selective rounds were necessary (minimum 3 rounds) in order to observe a significant shift of fluorescence. In that regard, we performed cytotoxicity tests on selected populations and we observed an inhibitory phenotype after only 2 rounds. In fact, our results show a significant drop in cell viability of 10 % after

Round 2 and a moderate drop of 5 % after Round 1 of selection. The observed evolution in phenotype is tending towards a selection of inhibitory scFvs but is still to be confirmed by other cell viability experiments.

Nowadays, the identification of the genotype that is responsible for a phenotype is a hot topic in medical research. Molecules with the potential to recognize, bind and modulate pathologic phenotypes play an important role for basic research, as well as in the generation of therapeutic agents. Two recent studies demonstrated that the apoptotic sensitivity in p53-mutated human cancer cells could be restored by microRNA (Herbert *et al.*, 2014) and a low molecular weight compound (Cui *et al.*, 2014). The intrabody-based phenotypic screening method described here aims to be an integrated approach in that domain. Intrabodies have demonstrated to summarize all the antibody properties within the cell. This includes enzyme inhibition (Paz *et al.*, 2005), breaking protein-protein (Griffin *et al.*, 2006) and protein-DNA interactions (Cohen *et al.*, 1998), re-activating mutant enzymes (Martineau *et al.*, 1998), targeting specific protein conformations, (Tanaka *et al.*, 2007) (Miller *et al.*, 2005), domains (Dauvillier *et al.*, 2002), and PTM (Cassimeris *et al.*, 2013); and inducing protein degradation (Caussinus *et al.*, 2012) (Butler and Messer, 2011). Additionally, by targeting them to specific cell compartments, intrabodies can re- or de-localize their target (Sibler *et al.*, 2003b) and block secretion (Böldicke, 2007). Furthermore, apart from proteins, antibodies are also able to recognize small chemicals usually referred as haptens, glycans and lipids (Rabu *et al.*, 2012) (Quintana *et al.*, 2012). The approach introduced here is different from all previous studies in that it used unbiased libraries that were not preselected against any known protein. As such, the strategy is analogous to classical forward genetic approaches except that it operates directly at the protein level.

In our model of allergy, we proved that the isolation of intrabodies as phenotype-actors can be realized without any previous knowledge of a target. Comparing our system to classical genetics, the intrabody plays the same role as mutations. Additionally, because the intrabody does not directly modify its target but modulates its function, it is also analogous to a chemical drug. The use of naïve intrabody library has been reported only by Lerner's group for the regulation of stem cell fate. By using combinatorial libraries, they found a single agonist intrabody against the alpha chains of integrins, which induced human stem cells to become dendritic cells (Yea *et al.*, 2013). Unlike typical approaches as shRNA and siRNA, in which the targets are known, nobody can predict the recognized target from an intrabody sequence. It is known that the immune system is able to raise antibodies against essentially any part of the surface of a protein (Jemmerson, 1987), and it remains to be demonstrated that this is also the case with intrabody libraries. Here we demonstrate the possibility

to construct a very efficient retroviral intrabody library and its use to identify cellular targets and specific modulators (inhibitors or activators). The approach described here represents a powerful tool not only to address the proteome diversity but also to study secondary messengers and metabolism in cells. As such, and compared to other large-scale approaches, this represents a straightforward path to the discovery of potential therapeutic molecules.

Annex 1: Materials and methods

1. Synthesis and optimization of the scFv library

1.1 Preparation of chemo-competent E. coli

Dut⁻ ung⁻ K12 CJ236 (NEB cat# E4141S) cells were streaked on LB agar + 35 µg/mL chloramphenicol (Cam). A single colony was picked out and cultured in 2 mL LB. This subculture was brought in 100 mL LB + 20 mM MgSO₄, and grown until OD₅₉₀ ≈ 0.5. After centrifugation, the pellet was resuspended in 20 mL ice cold TFB1 (30 mM KOAc, 100 mM RbCl, 10 mM CaCl₂, 50 mM MnCl₂, 15 % glycerol, pH 5.8). The cells were incubate on ice for 5' and centrifuged at 5000 rpm for 5' at 4°C. The pellet was resuspended in 4 mL cold TFB2 (10 mM MOPS, 75 mM CaCl₂, 10 mM RbCl, 15 % glycerol, pH 6.5) and incubated on ice for 1 hour before collecting in aliquots of 100 µL, stored at -80°C.

1.2 Preparation of uracil-containing single-stranded DNA (ssDNA)

10 ng of pCANTAB6-13R4-GFP vector was added to 50 µL of chemo-competent bacteria and incubated for 1 hour on ice. A heat shock of 45 seconds was performed at 37°C in a water bath. After 2' on ice, bacteria were diluted 10 times with LB and grown in a bacterial shaker for 30' at 37°C, 220 rpm. Bacteria were plated on ampicillin plates + 2 % glucose and incubate overnight, 37°C. Colonies were grown in 2 mL of 2xTY + Ampicillin + 2 % glucose. At an OD₆₀₀ ≈ 0.7 and 20X excess of helper phage KM13 was added. After an incubation of 1 hour at 37 °C, bacteria were centrifuged and the excess of KM13 was discarded. The pellet was resuspended in 2xTY + 2 % glucose, kanamycin, ampicillin and 0.25 µg/mL uridine and amplification was done overnight at 37 °C, 220 rpm. The day after, bacteria were pelleted and supernatants was collected, transferred to a fresh tube + 1/5 volume 20 % PEG-8000 and 2.5 M NaCl and incubated overnight on a 4 °C turning wheel. After a double centrifugation step for 10 ' at 10000 rpm, 2 °C, in order to remove all supernatants, the phage pellet was resuspended in 2 x 5 mL PBS and incubated for 10 ' at 2 °C. The insoluble matter was pelleted by centrifuging for 5' at 15000 rpm, 2°C. The supernatants was transferred to a 10 ml falcon and the yield was measured by subtracting OD₃₂₀ from OD₂₇₀. The single-stranded phage DNA was purified according to the QIAprep Spin M13 Protocol (#27704).

1.3 Annealing the library mutagenic primers to the ssDNA template

The primer/template molar ratio was set to 6:1. 20 µg dU-ssDNA was mixed with 75 pmol of each primer in presence annealing buffer. The annealing reaction is assembled on ice and directly transferred to a preheated thermal cycler at 90 °C for 2 ' for subsequent 10 seconds on ice to induce a cooling of the primer–template annealing. This allows the mixture to cool fast towards 45 °C, after 20' the temperature was dropped towards 20 °C for 10' followed by a hold on ice, ready for subsequent strand synthesis.

1.4 Synthesis of covalently closed circular scFv library: extension/ligation

The annealed complex of ssDNA and mutagenic primers was further completed by ligation for 4 hours at 20 °C in presence of dNTPs, T4 Ligase (Fermentas EL0011), T7 DNA polymerase (NEB M0274S) and ligase buffer. The purification was done by the NucleoSpin Plasmid Kit (Macherey-Nagel), according to the manufacturer instructions (except 1x AW wash and 2x A4 wash). The synthesis was evaluated on 1% agarose gel, alongside the ssDNA template. A successful reaction results in the complete conversion of single-stranded template to double-stranded DNA.

1.5 Transformation of bacteria with scFv library by electroporation

TG1 electro-competent cells (Lucigen) were thawed on ice for 15 '. The recovery medium (Lucigen) was thawed at RT. Electroporation cuvettes (0.2 cm gap), 1.5 mL tubes and the cuvette arm were chilled on ice. Bacteria were mixed carefully by tapping the tube and the scFv library dsDNA (10µg) was added. The mixture was briefly stirred and electroporated using the following conditions: 2500V, 25µF, 200Ω. The bacteria were immediately transferred to 11 mL recovery medium and incubated for 1 hour at 37°C. The diversity of the transformed bacteria was annotated by spreading dilutions on agar plates + ampicillin + 2% glucose, overnight at 37°C.

1.6 Purification of scFv library (MAXI prep)

The 11 mL culture of transformed bacteria was amplified in 500 mL 2xTY + 2 % glucose + ampicillin, overnight at 37 °C. The NucleoBond® Xtra for DNA purification kit (Macherey-Nagel) was used for prepping the plasmid DNA.

2. Cell cultures

2.1 Rat mast cell line RBL-2H3

RBL-2H3 cells (ATCC, Manassas, VA) were cultured in Dulbecco's modified Eagle's medium (DMEM) containing Glutamax ITM supplemented with 15% fetal calf serum (FCS) and antibiotics, at 37 °C in a humid incubator with 5 % CO₂. Adherent cells are passed three times per week. They are detached with 5' of trypsin at 37 °C, which is subsequently inactivated by the addition of two volumes of culture medium.

2.2 293T (or HEK-T)

These cells were maintained in culture in DMEM medium supplemented with 10% FCS and antibiotics, at 37 °C in a humid incubator with 5% CO₂. The hybridoma producing mouse monoclonal antibodies were maintained in culture in presence of DMEM supplemented with 10 % FCS, and

antibiotics at 37 °C in a wet incubator with 5% CO₂. The culture supernatants containing the antibodies were filtered, and stored at -20 °C. The concentration of antibody was determined by ELISA.

2.3 Colorectal cancer cell lines: HCT116, CaCo2 and DLD-1

Cell lines were cultured at 37 °C in a 5% CO₂ humidified atmosphere in Dulbecco's modified Eagle's medium (DMEM) containing Glutamax ITM, 10% (v/v) heat-inactivated fetal calf serum (FCS) and an antibiotic/antimycotic solution (Life Technologies).

3. Production and purification of antibody fragments

3.1 Production of antibody fragments from cytoplasmic extracts of bacteria

The antibody fragments were cloned into the pET23NN vector (modified from Novagen, allowing expression of a Myc-tag and a His-tag at the C-terminus of the scFv) between the *NcoI* and *NotI* sites.

3.2 Production of antibody fragments from bacterial periplasmic extracts

The antibody fragments were inserted into the vector pHEN2 at *NcoI* and *NotI* sites and HB2151 bacteria were transformed. After 16 hours of pre-culture at 16 °C in 2xTY containing 100 µg/mL ampicillin and 1 % glucose, the growth of bacteria was relaunched at 37 °C until an OD 600nm of 0.8 was reached. The induction was initiated by the addition of IPTG at a final concentration of 1 mM for 3 -4 h at 30 °C. Bacteria were then centrifuged at 3500 rpm for 20 ' and the pellets were lysed on ice for 15 ' in lysis buffer: 30 mM Tris pH 8, 20% sucrose, 1 mM EDTA, 1 mM PMSF. After centrifugation, 5 mM MgCl₂ and MgSO₄ were added to the periplasmic supernatants. The extracts were then stored at -20 °C.

3.3 Purification on Nickel resin

Antibody fragments were purified on nickel NTA resin manufactured by Qiagen. We followed the procedure of the Ni-NTA spin kit.

3.4 Purification on magnetic beads of nickel

For each immuno-precipitation scFvs were purified from bacterial extracts of cytoplasmic corresponding to 50 ml of culture with 20 µL of magnetic beads of nickel (Ademtech, Pessac, France), according to the specifications of the kit.

3.5 Production and immuno-precipitation of cell extracts

The cell layer was washed twice in warm RPMI. The activated cells were stimulated with 50-100 ng/mL of DNP-KLH in RPMI, 3-10' at 37 °C in the dark.

After washing with cold PBS containing phosphatase inhibitors (100 mM NaF, 5 mM orthovanadate), the cells were lysed for 15 ' on ice with lysis buffer containing PBS supplemented with: 0.5 % sodium deoxycholate, 1 % NP-40, 0.1 % SDS. After scratching the cells of the flasks, the cell lysates were clarified by centrifugation for 15' at 13,000 rpm at 4 °C. An aliquot of the supernatant containing the protein extracts was measured by means of a BC Assay kit (Uptima), the remaining lysates were supplemented with loading buffer (2 % SDS, 10 % glycerol, 2.5% β -mercaptoethanol, 0.01% bromophenol blue, 30 mM Tris pH6.8).

For immuno-precipitations, 3 mg protein lysates previously purified from cytoplasmic extracts of bacteria, containing the scFv, were extracted on magnetic beads of nickel; 2 hours at 4 °C. Before elution, 3 washes of 10' were performed with lysis buffer supplemented with 10 mM imidazole. Elution was performed by adding 500 mM imidazole, the eluate was then complemented with loading buffer and loaded on a gel for electrophoresis by SDS- PAGE.

4. HUSCIV expression in eukaryotic cells

4.1 Chemicals and enzymes

Chemicals were purchased from Sigma-Aldrich. Restriction enzymes and cloned Taq polymerase were from Fermentas. Phusion DNA polymerase was purchased from Finnzymes. Plasmid DNA, PCR and agarose-separated DNA were purified using Macherey2Nagel kits.

4.2 Plasmids

Plasmid pCMV/myc/cyto, obtained from Invitrogen (#V820-20), allows the expression of scFv genes from the strong CMV promoter in the cytoplasm of the cell. Plasmid pMSCVhygSN is derived from pMSCVhyg plasmid (Clontech) and is used for retroviral expression of the scFv as an N-terminal fusion with a c-myc and a His6 tag. First, a 207 bp PCR fragment was amplified from pAB1 plasmid (Martineau *et al.*, 1998a) using primers pelBbamHI2 (CCGCTGGATccTTACTC) and M13uni (AGGGTTTTCCAGTCACGACGTT).

4.3 Retrovirus production

Stable cell lines expressing the scFv library were obtained by retroviral gene transduction using scFv genes cloned in pMSCVhygSN-EGFP plasmid. Retroviral particles were produced in 293T by transient

co-transfection of gag/pol, env-VSV-G and the indicated viral pMSCV constructs. 2.10^6 cells were seeded on a 15 cm Petri dish the day before transfection. The pMSCV-derived vector (6 μ g), the packaging plasmid (2 μ g), the amphotropic envelope plasmid (2 μ g) and 1 ml of Jet Prime transfection buffer (Polyplus, Illkirch, France) and 20 μ L of jetPRIME[®] (Polyplus, Illkirch, France) transfection reagent were mixed and left at room temperature for 10 min. Then the mixture was added drop-wise to the cell culture medium. After 4 h, the medium was replaced with 20 mL of fresh medium. The supernatant containing the virus was collected 72 h later.

4.4 Retroviral transfection

10^6 HeLa and HCT116 cells were transduced in 15 cm Petri dishes with 6 mL of virus containing supernatant in the presence of 8 mg/mL polybrene (Sigma-Aldrich) for 24 hours. Successfully transfected cells were selected with 250 μ g/mL of Hygromycin.

4.5 Flow cytometry analysis and cell sorting

Single-cell suspension of cell lines expressing scFv-GFP fusions were analyzed by collecting at least 10 000 events/sample using an EPICS XLw cytometer (Beckman Coulter). Cells stably transduced with the retroviral scFv-GFP library were sorted for GFP expression with a FACS Aria cell sorter (Beckton Dickinson). One million of cells were resuspended in 1 mL of PBS.

4.6 Cell extracts, scFv precipitation and western blot analysis

The pelleted cells were lysed in SDS-PAGE sample buffer. For scFv precipitation, confluent cell cultures were lysed in a buffer containing 25 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% (w/v) Nonidet-P40, 1 mM Na_3VO_4 , 100 mM NaF, 1 mM phenylmethylsulphonyl fluoride and a protease inhibitor cocktail (Complete EDTA-free, Roche Applied Science). His-tagged proteins were captured from 3 mg of clarified lysates with 20 mL of magnetic beads (Ademtech, France). After incubation for 1 h at 48 °C under constant rotation, beads were washed twice according to the manufacture's protocol and the proteins were eluted.

Cells were cultured in Petri 15 cm flasks. For the recovery of the whole-cell protein content, transfected cells were harvested by trypsinization and washed twice in PBSi (PBS supplemented with complete protease inhibitor cocktail (Roche)). Cells were collected and sonicated (6 x 5 seconds with 20 % amplitude). After centrifugation at 16000 g for 30' at 4 °C, the soluble fractions were collected and re-centrifuged. The pellet was resuspended in PBSi and sonicated (3 x 5 seconds). The soluble fraction was quantified with the UPTIMA assay and used for nickel-precipitation as follows; 20 μ L Ademtech nickel beads were washed with binding buffer and 1.4 mg soluble extract was added and incubated on a turning wheel at 4 °C. After washing, the samples were stored at -20 °C in laemmli buffer (1x). Cell extracts and nickel-precipitates were analyzed by reducing SDS-PAGE (10% w/v polyacrylamide). Proteins were revealed by western blotting using a rabbit polyclonal serum against GFP (Santa Cruz, sc-8334), followed by a horseradish peroxidase-conjugated anti-rabbit secondary

antibody (1:15 000; Sigma-Aldrich) in PBS, 0.1 % Tween, 5 % milk. The signal was revealed using enhanced chemiluminescence and detected with a camera (GBOX Chemi, Syngene, Cambridge, UK).

4.7 Fluorescence microscopy

Cells were seeded on 12 mm glass coverslips in six-well plates. After 2 days, cells were fixed in 4 % PFA 10 min at RT. DNA was visualized using cell-permeant Hoechst 33342 dye at 5 mg/ml. The slides were washed with PBS, mounted in Mowiol and GFP-tagged scFv residing in cells were visualized with a fluorescence microscope (Zeiss) about 4 hours after fixation. To visualize scFv-GFP expression in live cells, coverslips were washed in PBS, laid on a drop of PBS on a glass slide and rapidly observed to minimize cell death.

5. Phenotypic targeting of mast cells and their role in allergy

5.1 Retrieving scFv sequences

Sequences were amplified out of mast cell chromosomal DNA (selective Round 7) with pMSCV.for/rev using Phusion DNA polymerase (GC buffer, 30 cycles). The sample was purified and diluted before applying the second PCR (Phusion GC, 35 cycles) with specific designed oligos (representing the 10 selected scFv sequence families, R_1 – R_10) and EGFPN1.rev. The DNA band of ~500 bp was sequenced with EGFPN1.rev by Eurofins MWG Operon and verified on both their variable domains.

5.2 Recloning in retroviral vector pMSCV-SN-EGFP

The vector/template pAB1-13R4 was amplified with M13rev-49.for and FR3.rev (Phusion, 35 cycles) in order to create a 454 bp fragment, which was PCR-assembled (M13rev-49, EGFPN1) to our sequences (5.4) in order to construct complete and functional scFvs. Those fragments are cloned into pMSCV-EGFP by digestion with *SfiI* and *NotI*.

5.3 Retrovirus production and transfection

Stable cell lines expressing the scFvs were obtained by retroviral gene transduction using scFv genes cloned in pMSCVhygSN-EGFP plasmid (5.5). Retroviral particles were produced in 293T by transient co-transfection of gag/pol, env-VSV-G and the indicated viral pMSCV constructs. 1.10^6 cells were seeded on a 10 cm Petri dish the day before transfection. The pMSCV-derived vector (3 µg), the packaging plasmid (1 µg), the amphotropic envelope plasmid (1 µg) and 500 µL of Jet Prime transfection buffer (Polyplus, Illkirch, France) and 10 µL of jetPRIME® (Polyplus, Illkirch, France) transfection reagent were mixed and left at room temperature for 10 min. Then the mixture was added drop-wise to the cell culture medium. After 5 h, the medium was replaced with 11 mL of fresh medium. The supernatant containing the virus was collected 72 h later.

10⁶ RBL cells were transduced in 10 cm Petri dishes with 2 mL of virus containing supernatant in the presence of 8 mg/mL polybrene (Sigma-Aldrich) for 24 hours. Successfully transfected cells were selected with 1 mg/mL of Hygromycin.

5.4 Analysis of stable transfectants

Measuring β -hexosaminidase release

The day before the experiment, 10⁵ cells per well were seeded in 96-well plates. 24h later, the adherent cells were activated by overnight addition of anti-DNP IgE followed by 50 to 200 ng/ml of DNP-KLH for 45 ' at 37 °C, as described above. After collecting the supernatant of each well (S1), the cells were lysed in lysis buffer (Tyrode buffer, 0.5% Triton, 50 μ g/mL aprotinin, 50 μ g/mL leupeptin, 50 μ g/mL pepstatin, 2 mM PMSF) for 20 ' on ice. The plate was then centrifuged for 5' at 2000 rpm in order to collect the supernatants and the corresponding cell lysates (S2).

The quantification of β -hexosaminidase is based on each 20 μ L of S1 and S2 supernatants, by the addition of 50 μ L of β -hexosaminidase substrate (p-dinitrophenyl-Nacetyl- β D-glucosaminidase, SIGMA) in a final concentration of 1.3 mg/mL for 1h30 at 37 °C. The substrate of β -hexosaminidase is freshly prepared or stored at -20 °C, in a 0.1 M citric acid solution, pH 4.5. The reaction was terminated by addition of 75 μ L of 0.4 M glycine, pH 10.7, and the intensity was evaluated by measuring the optical density at 405 nm. The percentage of β -hexosaminidase release was calculated as the ratio S1/ (S1 + S2) x100.

Measuring TNF- α release

The day before the experiment, 8.10⁵ cells per well were seeded in a 12-well culture plate. The next day, after washing in RPMI and Tyrode, cells were activated as described above, with 200 μ L per well of 50 ng/mL DNP-KLH for 2h at 37 °C. Supernatants are harvested and used for the quantification of TNF release, by means of an ELISA (BD) kit.

The cell layer was washed with cold PBS and lysed for 15 ' at 4 °C in PBS supplemented with 0.1% triton, and protease inhibitors (Complete Mini EDTA free tablet, Roche). In order to normalize the results, the amount of cellular protein was assayed using a BC Assay kit (Uptima).

High-throughput sequencing

Genomic DNA from 10⁶ cells (for the naïve library), or 5.10⁶ cells (for the selective rounds that followed) was extracted and amplified by PCR with primers adjacent to the VH and VL domains of the scFv. The DNA was prepared according to the procedures for sequencing. The analysis was performed at the MGX sequencing platform in Montpellier. The data obtained were analyzed using the SAM software.

6. Phenotypic targeting of *KRAS* mutated colorectal cancer cells

6.1 Retrovirus production

Stable cell lines expressing the scFv library were obtained by retroviral gene transduction using scFv genes cloned in pMSCVhygSN-EGFP plasmid. Retroviral particles were produced in 293T by transient co-transfection of gag/pol, env-VSV-G and the indicated viral pMSCV constructs. 2.10^6 cells were seeded on a 15 cm Petri dish the day before transfection. The pMSCV-derived vector (6 μ g), the packaging plasmid (2 μ g), the amphotropic envelope plasmid (2 μ g) and 1 ml of Jet Prime transfection buffer (Polyplus, Illkirch, France) and 20 μ L of jetPRIME[®] (Polyplus, Illkirch, France) transfection reagent were mixed and left at room temperature for 10 min. Then the mixture was added drop-wise to the cell culture medium. After 4 h, the medium was replaced with 20 mL of fresh medium. The supernatant containing the virus was collected 72 h later.

6.2 Viral titers

Viral titers were measured as follows: Cells of interest were cultured in 6-well plates at 0.5×10^6 cells/well. The freshly collected viral supernatants was diluted 10 to 10^6 times and added to the respective wells. After 10 days, clones were counted at the highest dilution possible. The clone number times the dilution factor represents the concentration of colony forming units (cfu). The same experiment was done with frozen supernatants in order to annotate a possible drop in virility.

6.3 Retro-transfection

Six million HCT116 cells were transduced in 15 cm Petri dishes with virus containing supernatant according to a ratio of 1:1, in the presence of 8 mg/mL polybrene (Sigma-Aldrich) for 24 h. Successfully transfected cells were selected with 250 mg/mL of Hygromycin.

6.4 FACS analysis of HCT116 stained with CMRA

Pass library-infected cells in (2x) 10 flasks (2.10^6 cells/Petri 15 cm). In the evening, add respectively 50 μ g/mL Cetuximab or Rituximab. Incubate for 3 days and pass the cells with 3.10^6 cells/flask. In the morning, add 2.5 μ M CMRA (in RPMI) to the cells, incubate for 30' and refresh medium. Incubate for 3 days. Wash petri flasks (10x 30.10^6 cells) with PBS and trypsinize. Pool the 10 flasks, take 50 % for FACS analysis and 50 % for DNA extraction. Resuspend in sterile tubes with 5 mL medium before cell sorting with FACS "Aria" (IRB, RIO imaging platform, Montpellier, $\pm 10^8$ cells). Use controls (CMRA⁻, CMRA⁺, wild type) to set-up the sorting parameters. For the first round, both Cetuximab and Rituximab treated cell pools were sorted. After recloning of the CMRA⁺ population, only Cetuximab-treated cells were sorted. All FACS files were analyzed and modified with FlowJo software.

6.5 Recloning of CMRA+ population

Genomic DNA from Cetuximab-treated CMRA⁺ pool was extracted with [QIAamp DNA Mini Kit](#). The scFv fragment was amplified out of 200 ng gDNA by PCR with primers pMSCV.for and pMSCV.rev using DreamTaq (Fermentas). The DNA band was purified with the DNA purification kit of Macherey-Nagel and quantified on agarose gel. The 1.7 kB scFv library fragments were cut with *SfiI* and *BglIII* in order to ligate it in pMSCVhygSB. The ligation was introduced in electro-competent bacteria, scFv DNA was amplified and purified as described previously.

6.6 Preparation of barcoded DNA fragments for Illumina sequencing

Genomic DNA of 100 million pooled library-transfected HCT116 cells treated for six days with respectively Cetuximab or Rituximab was extracted with Nucleobond CB500 (Macherey-Nagel). The scFv sequences were amplified as described above and a first digestion with *BstNI* was performed, which removes a contaminating sequence (parasite band). After that, a second PCR was performed in order to isolate the barcoded region. For this, primers egfp_BamHI.for and pMSCV.rev were used. This PCR introduced a BamHI site at the 5' end.

The DNA band (304 bp) was purified with the 'PCR and Gel Purification' kit of Macherey-Nagel. Next, the DNA fragment was digested with *BamHI* (FastDigest, Fermentas), creating sticky ends. The 8 bp fragment was eliminated with another purification step (Macherey-Nagel kit) and the 296 bp fragment was ligated to a dimer of 588 bp, containing two barcodes. After purification on gel the dimer was cut with *MlyI* (FastDigest, Fermentas), which creates a final DNA fragment of 356 bp containing two barcodes and with 5' and 3' ends ready for Illumina adaptor ligation.

Annex 2: Recloned intrabody family sequences

Table 1: Sequences of the most abundant retroviral clones from the 10 selected families. R_7 has a stop codon in the VL CDR3 loop and is thus truncated and expressed without the C-terminal eGFP tag. R_8 has the same VH than plasmid clone 5H4 and has been cloned as a single VH domain in retroviral vector.

Name	Selection	Family	VH CDR3	VL CDR3	Freq
R_1	Recloned	1	MDCVIGSYGYGIFDT	QSFVRNSTS	9.7E-4
R_2	Recloned	2	GKVLKKAEYSDWLDN	QQCSKFSLT	2.4E-3
R_3	Recloned	3	RSASCEH	EQYDTAPPYT	6.2E-3
R_4	Recloned	4	GEVGFYD	QQYFSQPFT	1.4E-3
R_5	Recloned	5	TLECSRCGDYGFDL	HQSNTPFT	1.6E-3
R_6	Recloned	6	DGLYARMYYNGSYY	QQYFSQPFT	2.9E-2
R_7	Recloned	7	ERRDDDGMAYSYQFDV	Q*	1.9E-2
R_8	Recloned	8	DGGLREGFDC	*	1.4E-3
R_9	Recloned	9	NPASKCVYLEHDFEK	QTCNCLTLV	1.3E-3
R_10	Recloned	10	PERSAYDY	QQYSSHPLT	2.6E-1

Annex 3: HUSCIV sequences

Table 2: Sequencing of randomly picked HUSCIV clones, before cloning in retroviral vector

	H1	H2	H3	L1	L2	L3
1	13R4	TCCATTAGTGGTAGTAGTAGATACATAGAT	13R4	13R4	13R4	13R4
2	13R4	13R4	13R4	13R4	13R4	13R4
3	AGTGAT	AACATTAGTGGTAGTAGTAGATATATAT	AGTAACTATAGT	13R4	GATGACAGTAAC	TATACAAACGGC
4	TATAGT	AGTATTTATGGTAGTAGTAGATATAAAGT	AGTAGTAACGGC	GGCTATTAT	AGTGACAGTTAT	TATACAAACAAC
5	TATTAT	AACATTAGTGGTAGTAGTAGATATAAAGT	AGTAACAGTAGTTTTGGTTAT	AGTTATAGT	AGTGACAGTAGT	TATACAAGTAAC
6	13R4	13R4	13R4	TATAACGAT	TATGACAGTTAT	TATACATATTAT
7	TATAAC	TATATTAGTGGTAGTAGTAGAAACATATAT	AGTAGTGATTATGGTGCC	AGTGATTAT	GATGACAGTAAC	GGCACAACATAT
8	13R4	13R4	13R4	13R4	13R4	13R4
9	TATTAT	AACATTGATGGTAGTAGTAGATATAAAC	AGTAACTATTATGGTGCC	13R4	13R4	TATACAAC TAAG
10	13R4	13R4	13R4	13R4	13R4	13R4
11	13R4	13R4	13R4	13R4	13R4	13R4
12	AGTAGT	AGTATTAGTGGTAGTAGTAGATATATAT	AGTTATTATTATTATTATGGCGGC	AATTATGAT	GATGACAGTTAT	TATACAACATAT
13	TATAAC	AACATTTATGGTAGTAGTAGATATAAAGT	AGTAGTTATAGTTTTGGTAAC	TATGATTAT	GAGGACAGTTAT	13R4
14	GGCGAT	AGTATTTATGGTAGTAGTAGAAACATATAT	AGTTATTATAACGGTGCC	TATTATTAT	AGTGACAGTTAT	AGTACAAGTAAC
15	13R4	13R4	TATAGTGCC	TATAGTAGT	AACGACAGTAAC	TATACAACATAT
16	13R4	13R4	13R4	13R4	13R4	13R4
17	13R4	13R4	AGTAATAATTATTTGGTGCC	13R4	13R4	TATACAACAAC
18	13R4	13R4	13R4	13R4	13R4	13R4
19	AACTAT	AACATTTATGGTAGTAGTAGATATAAAC	AGTAACAGTGGCGGTTAT	AACTATTAT	TATGACAGTAAC	TATACAGATGAT
20	TATGAT	TATATTGATGGTAGTAGTAGATATATAT	AGTTATTATTATGGTGCC	TATAACAGT	TATGACAGTTAT	AGTACATATGGC
21	13R4	13R4	13R4	13R4	13R4	13R4
22	TATTAT	TCCATTAGTGGTAGTAGTAGAAGTATAAAGT	AGTTATAGTTATGAT	AGTAACTAT	GATGACAGTGAT	CACACATATAGT
23	AGTAAC	TATATTTATGGTAGTAGTAGATATAAAC	AGTAGTAAC	TATTATAAC	GATGACAGTTAT	GATACAAGTAAC
24	13R4	13R4	AGTTATAGTAAC	13R4	13R4	13R4

25	13R4	13R4	13R4	13R4	13R4	13R4
26	13R4	13R4	13R4	GATTATTAT	13R4	13R4
27	13R4	13R4	GGCAACAAC	GATTATTAT	TATGACAGTAGT	TATACATATTAT
28	GATGGC	AACATTAGTGGTAGTAGTAGATATATAT	TATTATGGC	AGTTATAGT	TATGACAGTTAT	TATACAAGTTAT
29	13R4	13R4	13R4	13R4	AACGACAGTAGT	13R4
30	TATAAC	TATATTTATGGTAGTAGTAGATATAAGT	AGTAACGGCAGTGATAGTAACGGTAAC	AGTAACTAT	GATGACAGTTAT	TATACAGATGAT
31	13R4	TATATTGATGGTAGTAGTAGAAGTATAAAC	13R4	13R4	13R4	TATACAAACGGC
32	13R4	13R4	13R4	13R4	13R4	13R4
33	13R4	13R4	AGTAGTGATGGCGGTGGC	13R4	13R4	TATACATATTAT
34	13R4	13R4	13R4	TATGGCTAT	GAGGACAGTTAT	13R4
35	AGTAAC	AGTATTAACGGTAGTAGTAGAAGTATATAT	AGTTATTATTATGGTAAC	TATGATTAT	GATGACAGTTAT	AACACAAACTAT
36	AACGGC	TATATTTATGGTAGTAGTAGATATATAT	AGTAGTGCCAGT	TATTATTAT	TATGACAGTAAC	TATACAAGTTAT
37	13R4	13R4	13R4	13R4	13R4	13R4
38	13R4	13R4	13R4	13R4	13R4	13R4
39	13R4	13R4	AGTGATGATAAC	13R4	13R4	13R4
40	13R4	13R4	13R4	13R4	13R4	13R4
41	GATGGC	AGTATTTATGGTAGTAGTAGATATAAAC	13R4	TATAGTTAT	TATGACAGTAGT	AGTACATATAAC
42	13R4	13R4	AGTGGCGGCTAT	13R4	13R4	13R4
# total	42	42	42	42	42	42
# 13R4	25	23	19	22	21	20
%13R4	60%	55%	45%	52%	50%	48%

References

- Abraham, S.N., and St John, A.L. (2010). Mast cell-orchestrated immunity to pathogens. *Nat. Rev. Immunol.* *10*, 440–452.
- Absolom, D.R., and van Oss, C.J. (1986). The nature of the antigen-antibody bond and the factors affecting its association and dissociation. *CRC Crit. Rev. Immunol.* *6*, 1–46.
- Alvarez, R.D., Barnes, M.N., Gomez-Navarro, J., Wang, M., Strong, T.V., Arafat, W., Arani, R.B., Johnson, M.R., Roberts, B.L., Siegal, G.P., *et al.* (2000). A cancer gene therapy approach utilizing an anti-erbB-2 single-chain antibody-encoding adenovirus (AD21): a phase I trial. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* *6*, 3081–3087.
- Andreyev, H.J., Norman, A.R., Cunningham, D., Oates, J.R., and Clarke, P.A. (1998). Kirsten ras mutations in patients with colorectal cancer: the multicenter “RASCAL” study. *J. Natl. Cancer Inst.* *90*, 675–684.
- Barault, L., Veyrie, N., Jooste, V., Lecorre, D., Chapusot, C., Ferraz, J.-M., Lièvre, A., Cortet, M., Bouvier, A.-M., Rat, P., *et al.* (2008). Mutations in the RAS-MAPK, PI(3)K (phosphatidylinositol-3-OH kinase) signaling network correlate with poor survival in a population-based series of colon cancers. *Int. J. Cancer J. Int. Cancer* *122*, 2255–2259.
- Behr, T.M., Becker, W.S., Bair, H.J., Klein, M.W., Stühler, C.M., Cidlinsky, K.P., Wittekind, C.W., Scheele, J.R., and Wolf, F.G. (1995). Comparison of complete versus fragmented technetium-99m-labeled anti-CEA monoclonal antibodies for immunoscintigraphy in colorectal cancer. *J. Nucl. Med. Off. Publ. Soc. Nucl. Med.* *36*, 430–441.
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. (2003). GenBank. *Nucleic Acids Res.* *31*, 23–27.
- Benvenuti, S., Sartore-Bianchi, A., Di Nicolantonio, F., Zanon, C., Moroni, M., Veronese, S., Siena, S., and Bardelli, A. (2007). Oncogenic activation of the RAS/RAF signaling pathway impairs the response of metastatic colorectal cancers to anti-epidermal growth factor receptor antibody therapies. *Cancer Res.* *67*, 2643–2648.
- Bibeau, F., Lopez-Crapez, E., Di Fiore, F., Thezenas, S., Ychou, M., Blanchard, F., Lamy, A., Penault-Llorca, F., Frébourg, T., Michel, P., *et al.* (2009). Impact of FcγRIIIa-FcγRIIIa polymorphisms and KRAS mutations on the clinical outcome of patients with metastatic colorectal cancer treated with Cetuximab plus irinotecan. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* *27*, 1122–1129.
- Biocca, S., Neuberger, M.S., and Cattaneo, A. (1990). Expression and targeting of intracellular antibodies in mammalian cells. *EMBO J.* *9*, 101–108.
- Böldicke, T. (2007). Blocking translocation of cell surface molecules from the ER to the cell surface by intracellular antibodies targeted to the ER. *J. Cell. Mol. Med.* *11*, 54–70.
- Branch, A.D. (1998). A good antisense molecule is hard to find. *Trends Biochem. Sci.* *23*, 45–50.

- Braselmann, S., Taylor, V., Zhao, H., Wang, S., Sylvain, C., Baluom, M., Qu, K., Herlaar, E., Lau, A., Young, C., *et al.* (2006). R406, an orally available spleen tyrosine kinase inhibitor blocks fc receptor signaling and reduces immune complex-mediated inflammation. *J. Pharmacol. Exp. Ther.* *319*, 998–1008.
- Burt, R.W., Bishop, D.T., Lynch, H.T., Rozen, P., and Winawer, S.J. (1990). Risk and surveillance of individuals with heritable factors for colorectal cancer. WHO Collaborating Centre for the Prevention of Colorectal Cancer. *Bull. World Health Organ.* *68*, 655–665.
- Butler, D.C., and Messer, A. (2011). Bifunctional anti-huntingtin proteasome-directed intrabodies mediate efficient degradation of mutant huntingtin exon 1 protein fragments. *PLoS One* *6*, e29199.
- Cairns, M.J., Saravolac, E.G., and Sun, L.Q. (2002). Catalytic DNA: a novel tool for gene suppression. *Curr. Drug Targets* *3*, 269–279.
- Cassimeris, L., Guglielmi, L., Denis, V., Larroque, C., and Martineau, P. (2013). Specific *in vivo* labeling of tyrosinated α -tubulin and measurement of microtubule dynamics using a GFP tagged, cytoplasmically expressed recombinant antibody. *PLoS One* *8*, e59812.
- Castanotto, D., Li, J.R., Michienzi, A., Langlois, M.-A., Lee, N.S., Puymirat, J., and Rossi, J.J. (2002). Intracellular ribozyme applications. *Biochem. Soc. Trans.* *30*, 1140–1145.
- Cattaneo, A., and Biocca, S. (1999). The selection of intracellular antibodies. *Trends Biotechnol.* *17*, 115–121.
- Caussinus, E., Kanca, O., and Affolter, M. (2012). Fluorescent fusion protein knockout mediated by anti-GFP nanobody. *Nat. Struct. Mol. Biol.* *19*, 117–121.
- Cejas, P., López-Gómez, M., Aguayo, C., Madero, R., Moreno-Rubio, J., de Castro Carpeño, J., Beldaniesta, C., Barriuso, J., Moreno García, V., Díaz, E., *et al.* (2012). Analysis of the concordance in the EGFR pathway status between primary tumors and related metastases of colorectal cancer patients: implications for cancer therapy. *Curr. Cancer Drug Targets* *12*, 124–131.
- Chappell, W.H., Steelman, L.S., Long, J.M., Kempf, R.C., Abrams, S.L., Franklin, R.A., Bäsecke, J., Stivala, F., Donia, M., Fagone, P., *et al.* (2011). Ras/Raf/MEK/ERK and PI3K/PTEN/Akt/mTOR inhibitors: rationale and importance to inhibiting these pathways in human health. *Oncotarget* *2*, 135–164.
- De Chasse, B., Mikaelian, I., Mathieu, A.-L., Bickle, M., Olivier, D., Nègre, D., Cosset, F.-L., Rudkin, B.B., and Colas, P. (2007). An antiproliferative genetic screening identifies a peptide aptamer that targets calcineurin and up-regulates its activity. *Mol. Cell. Proteomics* *6*, 451–459.
- Chen, H., Ferbeyre, G., and Cedergren, R. (1997). Efficient hammerhead ribozyme and antisense RNA targeting in a slow ribosome *Escherichia coli* mutant. *Nat. Biotechnol.* *15*, 432–435.
- Cichowski, K., and Jänne, P.A. (2010). Drug discovery: inhibitors that activate. *Nature* *464*, 358–359.
- Citri, A., and Yarden, Y. (2006). EGF-ERBB signalling: towards the systems level. *Nat. Rev. Mol. Cell Biol.* *7*, 505–516.

- Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M.F., Kellis, M., Lindblad-Toh, K., and Lander, E.S. (2007). Distinguishing protein-coding and noncoding genes in the human genome. *Proc. Natl. Acad. Sci. U. S. A.* *104*, 19428–19433.
- Claverie, J.M. (2001). Gene number. What if there are only 30,000 human genes? *Science* *291*, 1255–1257.
- Cohen, S., and Milstein, C. (1967). Structure of antibody molecules. *Nature* *214*, 449–452 *passim*.
- Cohen, P.A., Mani, J.C., and Lane, D.P. (1998). Characterization of a new intrabody directed against the N-terminal region of human p53. *Oncogene* *17*, 2445–2456.
- Collis, A.V.J., Brouwer, A.P., and Martin, A.C.R. (2003). Analysis of the antigen combining site: correlations between length and sequence composition of the hypervariable loops and the nature of the antigen. *J. Mol. Biol.* *325*, 337–354.
- Dauvillier, S., Mérida, P., Visintin, M., Cattaneo, A., Bonnerot, C., and Dariavach, P. (2002). Intracellular single-chain variable fragments directed to the Src homology 2 domains of Syk partially inhibit Fc epsilon RI signaling in the RBL-2H3 cell line. *J. Immunol. Baltim. Md 1950* *169*, 2274–2283.
- David, M.P.C., Asprer, J.J.T., Ibane, J.S.A., Concepcion, G.P., and Padlan, E.A. (2007). A study of the structural correlates of affinity maturation: antibody affinity as a function of chemical interactions, structural plasticity and stability. *Mol. Immunol.* *44*, 1342–1351.
- Dawicki, W., and Marshall, J.S. (2007). New and emerging roles for mast cells in host defence. *Curr. Opin. Immunol.* *19*, 31–38.
- Delves, P.J., and Roitt, I.M. (2000). The immune system. First of two parts. *N. Engl. J. Med.* *343*, 37–49.
- Desiderio, A., Franconi, R., Lopez, M., Villani, M.E., Viti, F., Chiaraluce, R., Consalvi, V., Neri, D., and Benvenuto, E. (2001). A semi-synthetic repertoire of intrinsically stable antibody fragments derived from a single-framework scaffold. *J. Mol. Biol.* *310*, 603–615.
- Desplancq, D., Rinaldi, A.-S., Stoessel, A., Sibler, A.-P., Busso, D., Oulad-Abdelghani, M., Van Regenmortel, M.H., and Weiss, E. (2011). Single-chain Fv fragment antibodies selected from an intrabody library as effective mono- or bivalent reagents for *in vitro* protein detection. *J. Immunol. Methods* *369*, 42–50.
- Diaz, L.A., Jr, Williams, R.T., Wu, J., Kinde, I., Hecht, J.R., Berlin, J., Allen, B., Bozic, I., Reiter, J.G., Nowak, M.A., *et al.* (2012). The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* *486*, 537–540.
- Duff, S.E., Jeziorska, M., Rosa, D.D., Kumar, S., Haboubi, N., Sherlock, D., O'Dwyer, S.T., and Jayson, G.C. (2006). Vascular endothelial growth factors and receptors in colorectal cancer: implications for anti-angiogenic therapy. *Eur. J. Cancer Oxf. Engl.* *1990* *42*, 112–117.
- Duffy, A., and Kummar, S. (2009). Targeting mitogen-activated protein kinase kinase (MEK) in solid tumors. *Target. Oncol.* *4*, 267–273.

- Dvorak, H.F. (2002). Vascular permeability factor/vascular endothelial growth factor: a critical cytokine in tumor angiogenesis and a potential target for diagnosis and therapy. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* *20*, 4368–4380.
- Ellington, A.D., and Conrad, R. (1995). Aptamers as potential nucleic acid pharmaceuticals. *Biotechnol. Annu. Rev.* *1*, 185–214.
- Elsässer-Beile, U., Reischl, G., Wiehr, S., Bühler, P., Wolf, P., Alt, K., Shively, J., Judenhofer, M.S., Machulla, H.-J., and Pichler, B.J. (2009). PET imaging of prostate cancer xenografts with a highly specific antibody against the prostate-specific membrane antigen. *J. Nucl. Med. Off. Publ. Soc. Nucl. Med.* *50*, 606–611.
- Esteller, M. (2008). Epigenetics in cancer. *N. Engl. J. Med.* *358*, 1148–1159.
- Famulok, M. (1999). Oligonucleotide aptamers that recognize small molecules. *Curr. Opin. Struct. Biol.* *9*, 324–329.
- Fearon, E.R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell* *61*, 759–767.
- Forbes, S.A., Bhamra, G., Bamford, S., Dawson, E., Kok, C., Clements, J., Menzies, A., Teague, J.W., Futreal, P.A., and Stratton, M.R. (2008). The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr. Protoc. Hum. Genet.* Editor. Board Jonathan Haines *Chapter 10*, Unit 10.11.
- Galinsky, D.S.T., and Nechushtan, H. (2008). Mast cells and cancer--no longer just basic science. *Crit. Rev. Oncol. Hematol.* *68*, 115–130.
- Galli, S.J., and Tsai, M. (2010). Mast cells in allergy and infection: versatile effector and regulatory cells in innate and adaptive immunity. *Eur. J. Immunol.* *40*, 1843–1851.
- Galli, S.J., Maurer, M., and Lantz, C.S. (1999). Mast cells as sentinels of innate immunity. *Curr. Opin. Immunol.* *11*, 53–59.
- Garnett, M.J., and Marais, R. (2004). Guilty as charged: BRAF is a human oncogene. *Cancer Cell* *6*, 313–319.
- Garrido-Laguna, I., Hong, D.S., Janku, F., Nguyen, L.M., Falchook, G.S., Fu, S., Wheler, J.J., Luthra, R., Naing, A., Wang, X., *et al.* (2012). KRASness and PIK3CAness in patients with advanced colorectal cancer: outcome after treatment with early-phase trials with targeted pathway inhibitors. *PLoS One* *7*, e38033.
- Gattenlöhner, S., Etschmann, B., Kunzmann, V., Thalheimer, A., Hack, M., Kleber, G., Einsele, H., Germer, C., and Müller-Hermelink, H.-K. (2009). Concordance of KRAS/BRAF Mutation Status in Metastatic Colorectal Cancer before and after Anti-EGFR Therapy. *J. Oncol.* *2009*, 831626.
- Gilfillan, A.M., and Tkaczyk, C. (2006). Integrated signalling pathways for mast-cell activation. *Nat. Rev. Immunol.* *6*, 218–230.
- Gold, L., Polisky, B., Uhlenbeck, O., and Yarus, M. (1995). Diversity of oligonucleotide functions. *Annu. Rev. Biochem.* *64*, 763–797.

- Goldstein, N.S., and Armin, M. (2001). Epidermal growth factor receptor immunohistochemical reactivity in patients with American Joint Committee on Cancer Stage IV colon adenocarcinoma: implications for a standardized scoring system. *Cancer* 92, 1331–1346.
- Göringer, H.U., Homann, M., and Lorger, M. (2003). *In vitro* selection of high-affinity nucleic acid ligands to parasite target molecules. *Int. J. Parasitol.* 33, 1309–1317.
- Graus-Porta, D., Beerli, R.R., and Hynes, N.E. (1995). Single-chain antibody-mediated intracellular retention of ErbB-2 impairs Neu differentiation factor and epidermal growth factor signaling. *Mol. Cell. Biol.* 15, 1182–1191.
- Griffin, H., Elston, R., Jackson, D., Ansell, K., Coleman, M., Winter, G., and Doorbar, J. (2006). Inhibition of papillomavirus protein function in cervical cancer cells by intrabody targeting. *J. Mol. Biol.* 355, 360–378.
- Des Guetz, G., Uzzan, B., Nicolas, P., Cucherat, M., Morere, J.-F., Benamouzig, R., Breau, J.-L., and Perret, G.-Y. (2006). Microvessel density and VEGF expression are prognostic factors in colorectal cancer. Meta-analysis of the literature. *Br. J. Cancer* 94, 1823–1832.
- Guglielmi, L., Denis, V., Vezzio-Vié, N., Bec, N., Dariavach, P., Larroque, C., and Martineau, P. (2011). Selection for intrabody solubility in mammalian cells using GFP fusions. *Protein Eng. Des. Sel. PEDS* 24, 873–881.
- De Haard, H.J., van Neer, N., Reurs, A., Hufton, S.E., Roovers, R.C., Henderikx, P., de Bruïne, A.P., Arends, J.W., and Hoogenboom, H.R. (1999). A large non-immunized human Fab fragment phage library that permits rapid isolation and kinetic analysis of high affinity antibodies. *J. Biol. Chem.* 274, 18218–18230.
- Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hamers, C., Songa, E.B., Bendahman, N., and Hamers, R. (1993). Naturally occurring antibodies devoid of light chains. *Nature* 363, 446–448.
- Harriman, W., Völk, H., Defranoux, N., and Wabl, M. (1993). Immunoglobulin class switch recombination. *Annu. Rev. Immunol.* 11, 361–384.
- Harrison, P.M., Kumar, A., Lang, N., Snyder, M., and Gerstein, M. (2002). A question of size: the eukaryotic proteome and the problems in defining it. *Nucleic Acids Res.* 30, 1083–1090.
- Harwood, N.E., and Batista, F.D. (2010). Early events in B cell activation. *Annu. Rev. Immunol.* 28, 185–210.
- He, M., Menges, M., Groves, M.A., Corps, E., Liu, H., Brüggemann, M., and Taussig, M.J. (1999). Selection of a human anti-progesterone antibody fragment from a transgenic mouse library by ARM ribosome display. *J. Immunol. Methods* 231, 105–117.
- Heidorn, S.J., Milagre, C., Whittaker, S., Nourry, A., Niculescu-Duvas, I., Dhomen, N., Hussain, J., Reis-Filho, J.S., Springer, C.J., Pritchard, C., *et al.* (2010). Kinase-dead BRAF and oncogenic RAS cooperate to drive tumor progression through CRAF. *Cell* 140, 209–221.

Heinemann, V., Stintzing, S., Kirchner, T., Boeck, S., and Jung, A. (2009). Clinical relevance of EGFR- and KRAS-status in colorectal cancer patients treated with monoclonal antibodies directed against the EGFR. *Cancer Treat. Rev.* *35*, 262–271.

Heinemann, V., Douillard, J.Y., Ducreux, M., and Peeters, M. (2013). Targeted therapy in metastatic colorectal cancer -- an example of personalised medicine in action. *Cancer Treat. Rev.* *39*, 592–601.

Hoet, R.M., Cohen, E.H., Kent, R.B., Rookey, K., Schoonbroodt, S., Hogan, S., Rem, L., Frans, N., Daukandt, M., Pieters, H., *et al.* (2005). Generation of high-affinity human antibodies by combining donor-derived and synthetic complementarity-determining-region diversity. *Nat. Biotechnol.* *23*, 344–348.

Hudson, P.J. (1998). Recombinant antibody fragments. *Curr. Opin. Biotechnol.* *9*, 395–402.

Hyland, S., Beerli, R.R., Barbas, C.F., Hynes, N.E., and Wels, W. (2003). Generation and functional characterization of intracellular antibodies interacting with the kinase domain of human EGF receptor. *Oncogene* *22*, 1557–1567.

Jean, D., Bar-Eli, M., Huang, S., Xie, K., Rodrigues-Lima, F., Hermann, J., and Frade, R. (1996). A cysteine proteinase, which cleaves human C3, the third component of complement, is involved in tumorigenicity and metastasis of human melanoma. *Cancer Res.* *56*, 254–258.

Jemmerson, R. (1987). Multiple overlapping epitopes in the three antigenic regions of horse cytochrome c1. *J. Immunol. Baltim. Md 1950* *138*, 213–219.

Johnson, G., and Wu, T.T. (2000). Kabat database and its applications: 30 years after the first variability plot. *Nucleic Acids Res.* *28*, 214–218.

Jubb, A.M., Hurwitz, H.I., Bai, W., Holmgren, E.B., Tobin, P., Guerrero, A.S., Kabbinnar, F., Holden, S.N., Novotny, W.F., Frantz, G.D., *et al.* (2006). Impact of vascular endothelial growth factor-A expression, thrombospondin-2 expression, and microvessel density on the treatment effect of bevacizumab in metastatic colorectal cancer. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* *24*, 217–227.

Jung, Y.D., Mansfield, P.F., Akagi, M., Takeda, A., Liu, W., Bucana, C.D., Hicklin, D.J., and Ellis, L.M. (2002). Effects of combination anti-vascular endothelial growth factor receptor and anti-epidermal growth factor receptor therapies on the growth of gastric cancer in a nude mouse model. *Eur. J. Cancer Oxf. Engl.* *1990* *38*, 1133–1140.

Kabat, E.A., and Wu, T.T. (1991). Identical V region amino acid sequences and segments of sequences in antibodies of different specificities. Relative contributions of VH and VL genes, minigenes, and complementarity-determining regions to binding of antibody-combining sites. *J. Immunol. Baltim. Md 1950* *147*, 1709–1719.

Kalesnikoff, J., and Galli, S.J. (2008). New developments in mast cell biology. *Nat. Immunol.* *9*, 1215–1223.

Kim, D.-H., Longo, M., Han, Y., Lundberg, P., Cantin, E., and Rossi, J.J. (2004). Interferon induction by siRNAs and ssRNAs synthesized by phage polymerase. *Nat. Biotechnol.* *22*, 321–325.

- Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wölle, J., Plückthun, A., and Virnekäs, B. (2000). Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J. Mol. Biol.* *296*, 57–86.
- De Kruif, J., Boel, E., and Logtenberg, T. (1995). Selection and application of human single chain Fv antibody fragments from a semi-synthetic phage antibody display library with designed CDR3 regions. *J. Mol. Biol.* *248*, 97–105.
- Kunkel, T.A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. U. S. A.* *82*, 488–492.
- Kurreck, J. (2003). Antisense technologies. Improvement through novel chemical modifications. *Eur. J. Biochem. FEBS* *270*, 1628–1644.
- Laurent-Puig, P., Cayre, A., Manceau, G., Buc, E., Bachet, J.-B., Lecomte, T., Rougier, P., Lievre, A., Landi, B., Boige, V., *et al.* (2009). Analysis of PTEN, BRAF, and EGFR status in determining benefit from cetuximab therapy in wild-type KRAS metastatic colon cancer. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* *27*, 5924–5930.
- Lecerf, J.M., Shirley, T.L., Zhu, Q., Kazantsev, A., Amersdorfer, P., Housman, D.E., Messer, A., and Huston, J.S. (2001). Human single-chain Fv intrabodies counteract in situ huntingtin aggregation in cellular models of Huntington's disease. *Proc. Natl. Acad. Sci. U. S. A.* *98*, 4764–4769.
- Lemeh, C., and Arkenau, H.-T. (2012). Novel treatments for metastatic cutaneous melanoma and the management of emergent toxicities. *Clin. Med. Insights Oncol.* *6*, 53–66.
- Lener, M., Horn, I.R., Cardinale, A., Messina, S., Nielsen, U.B., Rybak, S.M., Hoogenboom, H.R., Cattaneo, A., and Biocca, S. (2000). Diverting a protein from its cellular location by intracellular antibodies. The case of p21Ras. *Eur. J. Biochem. FEBS* *267*, 1196–1205.
- Li, Z., Woo, C.J., Iglesias-Ussel, M.D., Ronai, D., and Scharff, M.D. (2004). The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev.* *18*, 1–11.
- Lin, W.-W., and Karin, M. (2007). A cytokine-mediated link between innate immunity, inflammation, and cancer. *J. Clin. Invest.* *117*, 1175–1183.
- Mandel, J.S., Bond, J.H., Church, T.R., Snover, D.C., Bradley, G.M., Schuman, L.M., and Ederer, F. (1993). Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. *N. Engl. J. Med.* *328*, 1365–1371.
- Marasco, W.A., Haseltine, W.A., and Chen, S.Y. (1993). Design, intracellular expression, and activity of a human anti-human immunodeficiency virus type 1 gp120 single-chain antibody. *Proc. Natl. Acad. Sci. U. S. A.* *90*, 7889–7893.
- Marisa, L., de Reyniès, A., Duval, A., Selves, J., Gaub, M.P., Vescovo, L., Etienne-Grimaldi, M.-C., Schiappa, R., Guenot, D., Ayadi, M., *et al.* (2013). Gene Expression Classification of Colon Cancer into Molecular Subtypes: Characterization, Validation, and Prognostic Value. *PLoS Med* *10*, e1001453.
- Martineau, P., Jones, P., and Winter, G. (1998a). Expression of an antibody fragment at high levels in the bacterial cytoplasm. *J. Mol. Biol.* *280*, 117–127.

- Mayr, L.M., and Bojanic, D. (2009). Novel trends in high-throughput screening. *Curr. Opin. Pharmacol.* *9*, 580–588.
- Mazuc, E., Villoutreix, B.O., Malbec, O., Roumier, T., Fleury, S., Leonetti, J.-P., Dombrowicz, D., Daëron, M., Martineau, P., and Dariavach, P. (2008). A novel druglike spleen tyrosine kinase binder prevents anaphylactic shock when administered orally. *J. Allergy Clin. Immunol.* *122*, 188–194, 194.e1–3.
- McGregor, D.P., Molloy, P.E., Cunningham, C., and Harris, W.J. (1994). Spontaneous assembly of bivalent single chain antibody fragments in *Escherichia coli*. *Mol. Immunol.* *31*, 219–226.
- De Melo, C.F.V., Gigeck, C.O., da Silva, J.N., Cardoso Smith, M. de A., de Araújo, R.M., Burbano, R.R., and Lima, E.M. (2014). Association of COX2 gene hypomethylation with intestinal type gastric cancer in samples of patients from northern Brazil. *Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med.* *35*, 1107–1111.
- Mendelsohn, J., and Baselga, J. (2003). Status of epidermal growth factor receptor antagonists in the biology and treatment of cancer. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* *21*, 2787–2799.
- Metcalfe, D.D., Baram, D., and Mekori, Y.A. (1997). Mast cells. *Physiol. Rev.* *77*, 1033–1079.
- Metz, M., Siebenhaar, F., and Maurer, M. (2008). Mast cell functions in the innate skin immune system. *Immunobiology* *213*, 251–260.
- Mhashilkar, A.M., Bagley, J., Chen, S.Y., Szilvay, A.M., Helland, D.G., and Marasco, W.A. (1995). Inhibition of HIV-1 Tat-mediated LTR transactivation and HIV-1 infection by anti-Tat single chain intrabodies. *EMBO J.* *14*, 1542–1551.
- Mhashilkar, A.M., Doebis, C., Seifert, M., Busch, A., Zani, C., Soo Hoo, J., Nagy, M., Ritter, T., Volk, H.-D., and Marasco, W.A. (2002). Intrabody-mediated phenotypic knockout of major histocompatibility complex class I expression in human and monkey cell lines and in primary human keratinocytes. *Gene Ther.* *9*, 307–319.
- Miller, T.W., Zhou, C., Gines, S., MacDonald, M.E., Mazarakis, N.D., Bates, G.P., Huston, J.S., and Messer, A. (2005). A human single-chain Fv intrabody preferentially targets amino-terminal Huntingtin's fragments in striatal models of Huntington's disease. *Neurobiol. Dis.* *19*, 47–56.
- Misale, S., Yaeger, R., Hobor, S., Scala, E., Janakiraman, M., Liska, D., Valtorta, E., Schiavo, R., Buscarino, M., Siravegna, G., *et al.* (2012). Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* *486*, 532–536.
- Miyazaki, M., Furuya, T., Shiraki, A., Sato, T., Oga, A., and Sasaki, K. (1999). The relationship of DNA ploidy to chromosomal instability in primary human colorectal cancers. *Cancer Res.* *59*, 5283–5285.
- Montagut, C., Dalmases, A., Bellosillo, B., Crespo, M., Pairet, S., Iglesias, M., Salido, M., Gallen, M., Marsters, S., Tsai, S.P., *et al.* (2012). Identification of a mutation in the extracellular domain of the Epidermal Growth Factor Receptor conferring cetuximab resistance in colorectal cancer. *Nat. Med.* *18*, 221–223.

- Nieba, L., Honegger, A., Krebber, C., and Plückthun, A. (1997). Disrupting the hydrophobic patches at the antibody variable/constant domain interface: improved *in vivo* folding and physical characterization of an engineered scFv fragment. *Protein Eng.* *10*, 435–444.
- Niu, G., Wright, K.L., Huang, M., Song, L., Haura, E., Turkson, J., Zhang, S., Wang, T., Sinibaldi, D., Coppola, D., *et al.* (2002). Constitutive Stat3 activity up-regulates VEGF expression and tumor angiogenesis. *Oncogene* *21*, 2000–2008.
- Nizak, C., Monier, S., del Nery, E., Moutel, S., Goud, B., and Perez, F. (2003). Recombinant antibodies to the small GTPase Rab6 as conformation sensors. *Science* *300*, 984–987.
- Noelle, R., and Snow, E.C. (1992). T helper cells. *Curr. Opin. Immunol.* *4*, 333–337.
- Noli, C., and Miolo, A. (2001). The mast cell in wound healing. *Vet. Dermatol.* *12*, 303–313.
- Norrby, K. (2006). *In vivo* models of angiogenesis. *J. Cell. Mol. Med.* *10*, 588–612.
- Ochs, H.D., and Wedgwood, R.J. (1987). IgG subclass deficiencies. *Annu. Rev. Med.* *38*, 325–340.
- Ohage, E.C., Wirtz, P., Barnikow, J., and Steipe, B. (1999). Intrabody construction and expression. II. A synthetic catalytic Fv fragment. *J. Mol. Biol.* *291*, 1129–1134.
- Padlan, E.A. (1994). Anatomy of the antibody molecule. *Mol. Immunol.* *31*, 169–217.
- Padlan, E.A., and Kabat, E.A. (1991). Modeling of antibody combining sites. *Methods Enzymol.* *203*, 3–21.
- Padlan, E.A., Abergel, C., and Tipper, J.P. (1995). Identification of specificity-determining residues in antibodies. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* *9*, 133–139.
- Palm, N.W., and Medzhitov, R. (2007). Not so fast: adaptive suppression of innate immunity. *Nat. Med.* *13*, 1142–1144.
- Pan, W., Craven, R.C., Qiu, Q., Wilson, C.B., Wills, J.W., Golovine, S., and Wang, J.F. (1995). Isolation of virus-neutralizing RNAs from a large pool of random sequences. *Proc. Natl. Acad. Sci. U. S. A.* *92*, 11509–11513.
- Parker, D.C. (1993). T cell-dependent B cell activation. *Annu. Rev. Immunol.* *11*, 331–360.
- Paz, K., Brennan, L.A., Iacolina, M., Doody, J., Hadari, Y.R., and Zhu, Z. (2005). Human single-domain neutralizing intrabodies directed against Etk kinase: a novel approach to impair cellular transformation. *Mol. Cancer Ther.* *4*, 1801–1809.
- Pentheroudakis, G., Kotoula, V., De Roock, W., Kouvatsas, G., Papakostas, P., Makatsoris, T., Papamichael, D., Xanthakis, I., Sgouros, J., Televantou, D., *et al.* (2013). Biomarkers of benefit from cetuximab-based therapy in metastatic colorectal cancer: interaction of EGFR ligand expression with RAS/RAF, PIK3CA genotypes. *BMC Cancer* *13*, 49.
- Philibert, P., Stoessel, A., Wang, W., Sibler, A.-P., Bec, N., Larroque, C., Saven, J.G., Courtête, J., Weiss, E., and Martineau, P. (2007). A focused antibody library for selecting scFvs expressed at high levels in the cytoplasm. *BMC Biotechnol.* *7*, 81.

- Phillips, J., Artsaenko, O., Fiedler, U., Horstmann, C., Mock, H.P., Müntz, K., and Conrad, U. (1997). Seed-specific immunomodulation of abscisic acid activity induces a developmental switch. *EMBO J.* *16*, 4489–4496.
- Porter R.R. (1963). Chemical structure of gamma-globulin and antibodies. *Br. Med. Bull.* *19*, 197–201.
- Qiao, H., Andrade, M.V., Lisboa, F.A., Morgan, K., and Beaven, M.A. (2006). FcεR1 and toll-like receptors mediate synergistic signals to markedly augment production of inflammatory cytokines in murine mast cells. *Blood* *107*, 610–618.
- Quintana, F.J., Yeste, A., Weiner, H.L., and Covacu, R. (2012). Lipids and lipid-reactive antibodies as biomarkers for multiple sclerosis. *J. Neuroimmunol.* *248*, 53–57.
- Rabu, C., McIntosh, R., Jurasova, Z., and Durrant, L. (2012). Glycans as targets for therapeutic antitumor antibodies. *Future Oncol. Lond. Engl.* *8*, 943–960.
- Reth, M. (1992). Antigen receptors on B lymphocytes. *Annu. Rev. Immunol.* *10*, 97–121.
- Ribatti, D., and Crivellato, E. (2012). Mast cells, angiogenesis, and tumour growth. *Biochim. Biophys. Acta* *1822*, 2–8.
- Richardson, J.H., Sodroski, J.G., Waldmann, T.A., and Marasco, W.A. (1995). Phenotypic knockout of the high-affinity human interleukin 2 receptor by intracellular single-chain antibodies against the alpha subunit of the receptor. *Proc. Natl. Acad. Sci. U. S. A.* *92*, 3137–3141.
- Rinaldi, A.-S., Freund, G., Desplancq, D., Sibling, A.-P., Baltzinger, M., Rochel, N., Mély, Y., Didier, P., and Weiss, E. (2013). The use of fluorescent intrabodies to detect endogenous gankyrin in living cancer cells. *Exp. Cell Res.* *319*, 838–849.
- De Roock, W., Piessevaux, H., De Schutter, J., Janssens, M., De Hertogh, G., Personeni, N., Biesmans, B., Van Laethem, J.-L., Peeters, M., Humblet, Y., *et al.* (2008). KRAS wild-type state predicts survival and is associated to early radiological response in metastatic colorectal cancer treated with cetuximab. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. ESMO* *19*, 508–515.
- De Roock, W., De Vriendt, V., Normanno, N., Ciardiello, F., and Tejpar, S. (2011). KRAS, BRAF, PIK3CA, and PTEN mutations: implications for targeted therapies in metastatic colorectal cancer. *Lancet Oncol.* *12*, 594–603.
- Rossi, J.J. (1995). Controlled, targeted, intracellular expression of ribozymes: progress and problems. *Trends Biotechnol.* *13*, 301–306.
- Samowitz, W.S. (2007). The CpG island methylator phenotype in colorectal cancer. *J. Mol. Diagn. JMD* *9*, 281–283.
- Sanz, L., Cuesta, A.M., Compte, M., and Alvarez-Vallina, L. (2005). Antibody engineering: facing new challenges in cancer therapy. *Acta Pharmacol. Sin.* *26*, 641–648.
- Scaltriti, M., and Baselga, J. (2006). The epidermal growth factor receptor pathway: a model for targeted therapy. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* *12*, 5268–5272.

- Shashikant, C.S., and Ruddle, F.H. (2003). Impact of transgenic technologies on functional genomics. *Curr. Issues Mol. Biol.* 5, 75–98.
- Sibler, A.-P., Nordhammer, A., Masson, M., Martineau, P., Travé, G., and Weiss, E. (2003). Nucleocytoplasmic shuttling of antigen in mammalian cells conferred by a soluble versus insoluble single-chain antibody fragment equipped with import/export signals. *Exp. Cell Res.* 286, 276–287.
- Sibler, A.-P., Courtête, J., Muller, C.D., Zeder-Lutz, G., and Weiss, E. (2005). Extended half-life upon binding of destabilized intrabodies allows specific detection of antigen in mammalian cells. *FEBS J.* 272, 2878–2891.
- Skerra, A., and Plückthun, A. (1988). Assembly of a functional immunoglobulin Fv fragment in *Escherichia coli*. *Science* 240, 1038–1041.
- Spano, J.-P., Lagorce, C., Atlan, D., Milano, G., Domont, J., Benamouzig, R., Attar, A., Benichou, J., Martin, A., Morere, J.-F., *et al.* (2005). Impact of EGFR expression on colorectal cancer patient prognosis and survival. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. ESMO* 16, 102–108.
- Steinberger, P., Andris-Widhopf, J., Bühler, B., Torbett, B.E., and Barbas, C.F., 3rd (2000). Functional deletion of the CCR5 receptor by intracellular immunization produces cells that are refractory to CCR5-dependent HIV-1 infection and cell fusion. *Proc. Natl. Acad. Sci. U. S. A.* 97, 805–810.
- Stewart, M.W. (2014). Anti-VEGF Therapy for Diabetic Macular Edema. *Curr. Diab. Rep.* 14, 510.
- Stockwell, B.R. (2000). Chemical genetics: ligand-based discovery of gene function. *Nat. Rev. Genet.* 1, 116–125.
- Strube, R.W., and Chen, S.-Y. (2002). Characterization of anti-cyclin E single-chain Fv antibodies and intrabodies in breast cancer cells: enhanced intracellular stability of novel sFv-F(c) intrabodies. *J. Immunol. Methods* 263, 149–167.
- Tanaka, T., and Rabbitts, T.H. (2003). Intrabodies based on intracellular capture frameworks that bind the RAS protein with high affinity and impair oncogenic transformation. *EMBO J.* 22, 1025–1035.
- Tanaka, T., and Rabbitts, T.H. (2012). Intracellular antibody capture (IAC) methods for single domain antibodies. *Methods Mol. Biol. Clifton NJ* 911, 151–173.
- Tanaka, T., Williams, R.L., and Rabbitts, T.H. (2007). Tumour prevention by a single antibody domain targeting the interaction of signal transduction proteins with RAS. *EMBO J.* 26, 3250–3259.
- Tijsterman, M., and Plasterk, R.H.A. (2004). Dicers at RISC; the mechanism of RNAi. *Cell* 117, 1–3.
- Tol, J., and Punt, C.J.A. (2010). Monoclonal antibodies in the treatment of metastatic colorectal cancer: a review. *Clin. Ther.* 32, 437–453.
- Tol, J., Nagtegaal, I.D., and Punt, C.J.A. (2009). BRAF mutation in metastatic colorectal cancer. *N. Engl. J. Med.* 361, 98–99.
- Tolstrup, A.B., Duch, M., Dalum, I., Pedersen, F.S., and Mouritsen, S. (2001). Functional screening of a retroviral peptide library for MHC class I presentation. *Gene* 263, 77–84.

- Tse, E., Lobato, M.N., Forster, A., Tanaka, T., Chung, G.T.Y., and Rabbitts, T.H. (2002). Intracellular antibody capture technology: application to selection of intracellular antibodies recognising the BCR-ABL oncogenic protein. *J. Mol. Biol.* *317*, 85–94.
- Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* *249*, 505–510.
- Uhlmann, E., Peyman, A., Ryte, A., Schmidt, A., and Buddecke, E. (2000). Use of minimally modified antisense oligonucleotides for specific inhibition of gene expression. *Methods Enzymol.* *313*, 268–284.
- Villoutreix, B.O., Laconde, G., Lagorce, D., Martineau, P., Miteva, M.A., and Dariavach, P. (2011). Tyrosine kinase syk non-enzymatic inhibitors and potential anti-allergic drug-like compounds discovered by virtual and *in vitro* screening. *PLoS One* *6*, e21117.
- Visintin, M., Tse, E., Axelson, H., Rabbitts, T.H., and Cattaneo, A. (1999). Selection of antibodies for intracellular function using a two-hybrid *in vivo* system. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 11723–11728.
- Visintin, M., Settanni, G., Maritan, A., Graziosi, S., Marks, J.D., and Cattaneo, A. (2002). The intracellular antibody capture technology (IACT): towards a consensus sequence for intracellular antibodies. *J. Mol. Biol.* *317*, 73–83.
- Waldo, G.S., Standish, B.M., Berendzen, J., and Terwilliger, T.C. (1999). Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* *17*, 691–695.
- Ward, A.F., Braun, B.S., and Shannon, K.M. (2012). Targeting oncogenic Ras signaling in hematologic malignancies. *Blood* *120*, 3397–3406.
- Ward, E.S., Güssow, D., Griffiths, A.D., Jones, P.T., and Winter, G. (1989). Binding activities of a repertoire of single immunoglobulin variable domains secreted from *Escherichia coli*. *Nature* *341*, 544–546.
- Willis, J.R., Briney, B.S., DeLuca, S.L., Crowe, J.E., and Meiler, J. (2013). Human germline antibody gene segments encode polyspecific antibodies. *PLoS Comput. Biol.* *9*, e1003045.
- Wingo, P.A., Tong, T., and Bolden, S. (1995). Cancer statistics, 1995. *CA. Cancer J. Clin.* *45*, 8–30.
- Winter, G., and Milstein, C. (1991). Man-made antibodies. *Nature* *349*, 293–299.
- Wörn, A., and Plückthun, A. (1998). An intrinsically stable antibody scFv fragment can tolerate the loss of both disulfide bonds and fold correctly. *FEBS Lett.* *427*, 357–361.
- Wörn, A., and Plückthun, A. (2001). Stability engineering of antibody single-chain Fv fragments. *J. Mol. Biol.* *305*, 989–1010.
- Xie, J., Yea, K., Zhang, H., Moldt, B., He, L., Zhu, J., and Lerner, R.A. (2014). Prevention of cell death by antibodies selected from intracellular combinatorial libraries. *Chem. Biol.* *21*, 274–283.
- Xu, H., Zhang, P., Liu, L., and Lee, M.Y. (2001). A novel PCNA-binding motif identified by the panning of a random peptide display library. *Biochemistry (Mosc.)* *40*, 4512–4520.

Yang, Z.-Y., Shen, W.-X., Hu, X.-F., Zheng, D.-Y., Wu, X.-Y., Huang, Y.-F., Chen, J.-Z., Mao, C., and Tang, J.-L. (2012). EGFR gene copy number as a predictive biomarker for the treatment of metastatic colorectal cancer with anti-EGFR monoclonal antibodies: a meta-analysis. *J. Hematol. Oncol.* *J Hematol Oncol* *5*, 52.

Yea, K., Zhang, H., Xie, J., Jones, T.M., Yang, G., Song, B.D., and Lerner, R.A. (2013). Converting stem cells to dendritic cells by agonist antibodies from unbiased morphogenic selections. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 14966–14971.

Zhang, J., Billingsley, M.L., Kincaid, R.L., and Siraganian, R.P. (2000). Phosphorylation of Syk activation loop tyrosines is essential for Syk function. An *in vivo* study using a specific anti-Syk activation loop phosphotyrosine antibody. *J. Biol. Chem.* *275*, 35442–35447.



In-Cell Intrabody Selection from a Diverse Human Library Identifies C12orf4 Protein as a New Player in Rodent Mast Cell Degranulation

Elsa Mazuc^{1,2,3,4,9}, Laurence Guglielmi^{1,2,3,4,9}, Nicole Bec^{1,2,3,4}, Vincent Perez^{1,2,3,4}, Chang S. Hahn⁵, Caroline Mollevi⁴, Hugues Parrinello⁶, Jean-Pierre Desvignes⁶, Christian Larroque^{1,2,3,4}, Ray Jupp⁵, Piona Dariavach^{1,2,3,4,7,*†}, Pierre Martineau^{1,2,3,4,*†}

1 IRCM, Institut de Recherche en Cancérologie de Montpellier, Montpellier, France, **2** INSERM, U896, Montpellier, France, **3** Université Montpellier1, Montpellier, France, **4** ICM, Institut régional du Cancer Montpellier, Montpellier, France, **5** Sanofi-Aventis, Bridgewater, New Jersey, United States of America, **6** MGX-Montpellier GenomiX, Institut de Génomique Fonctionnelle, Montpellier, France, **7** Université Montpellier2, Montpellier, France

Abstract

The high specificity of antibodies for their antigen allows a fine discrimination of target conformations and post-translational modifications, making antibodies the first choice tool to interrogate the proteome. We describe here an approach based on a large-scale intracellular expression and selection of antibody fragments in eukaryotic cells, so-called intrabodies, and the subsequent identification of their natural target within living cell. Starting from a phenotypic trait, this integrated system allows the identification of new therapeutic targets together with their companion inhibitory intrabody. We applied this system in a model of allergy and inflammation. We first cloned a large and highly diverse intrabody library both in a plasmid and a retroviral eukaryotic expression vector. After transfection in the RBL-2H3 rat basophilic leukemia cell line, we performed seven rounds of selection to isolate cells displaying a defect in FcεRI-induced degranulation. We used high throughput sequencing to identify intrabody sequences enriched during the course of selection. Only one intrabody was common to both plasmid and retroviral selections, and was used to capture and identify its target from cell extracts. Mass spectrometry analysis identified protein RGD1311164 (C12orf4), with no previously described function. Our data demonstrate that RGD1311164 is a cytoplasmic protein implicated in the early signaling events following FcεRI-induced cell activation. This work illustrates the strength of the intrabody-based in-cell selection, which allowed the identification of a new player in mast cell activation together with its specific inhibitor intrabody.

Citation: Mazuc E, Guglielmi L, Bec N, Perez V, Hahn CS, et al. (2014) In-Cell Intrabody Selection from a Diverse Human Library Identifies C12orf4 Protein as a New Player in Rodent Mast Cell Degranulation. PLoS ONE 9(8): e104998. doi:10.1371/journal.pone.0104998

Editor: Stefan Dübel, Technical University of Braunschweig, Germany

Received: May 24, 2014; **Accepted:** July 14, 2014; **Published:** August 14, 2014

Copyright: © 2014 Mazuc et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the Sanofi-Aventis collaboration agreement N° L07053, and Labex MabiImprove. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have read the journal's policy and the authors of this manuscript have the following competing interests: CSH and RJ are employed by Sanofi-Aventis. This does not alter their adherence to PLOS ONE policies on sharing data and materials.

* Email: piona.dariavach@inserm.fr (PD); pierre.martineau@inserm.fr (PM)

† These authors contributed equally to this work.

‡ These authors also contributed equally to this work.

Introduction

Mast cells and basophils are key effector cells in IgE-associated immediate hypersensitivity and allergic disorders. Upon FcεRI crosslinking initiated by the binding of antigen-IgE complexes, cell activation results in downstream events that lead to the secretion of three classes of mediators: (a) the extracellular release of preformed mediators stored in cell cytoplasmic granules, by a process called degranulation; (b) the de novo synthesis of proinflammatory lipid mediators; and (c) the synthesis and secretion of many growth factors, cytokines, and chemokines. This IgE-dependent release of mediators begins within minutes of antigen challenge and leads to certain acute allergic reactions such as anaphylaxis and acute attacks of atopic asthma [1].

The majority of drugs currently used to treat allergic disorders target only a single mediator released by mast cells. Examples include antihistamine H1 receptor antagonists, leukotriene modifiers, and steroids that predominantly inhibit mast-cell mediator production. More recently, protein therapies have permitted alternative approaches in addition to drug therapies. In this respect, an important treatment for allergic conditions is the recombinant humanized IgG monoclonal antibody Omalizumab, which binds selectively to human IgE and inhibit the production and release of all mast cell mediators by antagonizing IgE action. Although this biologic is highly effective, it is difficult and expensive to manufacture and administer.

An alternative that has gained significant attention in recent years is to target key enzymes involved in the signal transduction pathways initiated following FcεRI crosslinking. Mast cell activa-

tion results from the transient perturbation of an active balance between positive and negative signals that is consequent to engagement of membrane receptors. Classically, kinases and phosphatases have been viewed as the effectors of positive and negative signals, respectively. FcεRI mainly trigger positive signals by recruiting tyrosine kinases and signalosomes into which signaling molecules assemble [2].

In the past decade, one of the compelling targets for the treatment of allergic and autoimmune disorders was the Spleen tyrosine kinase (Syk), a key mediator of immunoreceptor signaling [3]. Many pharmaceutical companies as well as academic institutions have been involved in the development of small-molecule inhibitors of Syk that target the conserved ATP binding site within the catalytic domain of the kinase. But due to the similarities of the ATP pocket structures among different kinases, the ATP-binding site inhibitors of Syk affect multiple tyrosine kinases and have off-target effects that lead to undesirable side effects [4]. For these reasons, clinical trials using systemic modes of administration of Syk inhibitors were abandoned in favor of local modes of administration. Examples are the compound R112, the first Syk inhibitor to enter clinical studies developed by Rigel as an intranasal administration for seasonal allergic rhinitis [5] and R343, an inhaled formulation for the treatment of allergic asthma (Pfizer) [6].

In our previous studies, we devised an approach to identify protein-protein interaction and allosteric inhibitors of Syk instead of targeting its catalytic site. Our goal was to improve the selectivity and the safety profiles of Syk inhibitor drug candidates by selecting drugs targeting the SH2 domains of Syk. To achieve this, we developed an antibody displacement assay to convert an intrabody directed against the SH2 domain of Syk into chemical drugs [7,8]. The isolated molecules recapitulated the intrabody effects in cell cultures and were able to block the anaphylactic shock when administered orally in animal models [7]. This led to the identification of several scaffolds as potential starting points for the development of new classes of non-enzymatic inhibitors of Syk with minimal off-target effects [7,8].

The anti-Syk inhibitory intrabody used in the above studies was selected from a two-step process: a) *in vitro* screen of a phage display library against a recombinant Syk protein; and b) intracellular expression of the isolated antibody fragments in mammalian cells to test their inhibitory potential [9]. However, all targets may not be identified beforehand and it is tempting to envision a direct selection of intrabodies in cells. The use of an intrabody library as a target discovery platform has been suggested almost 20 years ago by Dr. A. Cattaneo and co-workers [10,11]. In a pioneering experiment, they demonstrated the rescue of an antiviral neutralizing intrabody diluted within a polyclonal repertoire. Although this was obtained in a model system with a limited diversity, this paved the way to a direct selection of a diverse repertoire of antibody fragments based on a selectable cell phenotype [12].

We report here the first application of such a strategy for the identification of new therapeutic targets in the field of allergy and inflammation. The Intrabody-based Phenotypic Screen (IBPheS) is based on the intracellular expression of a highly diverse antibody fragment library in eukaryotic cells and the selection of antibody fragments associated with the desired phenotype. Because this method is based on the intrabody-target interaction, it results in the co-selection of a target with its companion intrabody. This ensures that the fished targets are accessible to in-cell modulation (inhibition or activation) and thus may represent a “druggable” class of proteins. In the complex biological system of FcεRI-induced cell activation, the IBPheS approach led to the discovery

of a protein of yet unknown function implicated in mast cell degranulation.

Materials and Methods

Reagents and Antibodies

Antibodies were obtained from Santa Cruz Biotechnology (Santa Cruz, CA, USA), with the exception of anti-p44/42 MAP Kinase antibody and the following phospho-specific antibodies which were obtained from Cell Signaling Technologies (Danvers, MA, USA): Phospho Syk (Tyr525/526); Phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204); Phospho-SAPK/JNK (Thr183/Tyr185); Phospho-p38 MAP Kinase (Thr180/Tyr182); Phospho-Src Family (Tyr416); Phospho-NFκB (Ser536); Phospho-Gab2 (Tyr452); Phospho-PLCγ2 (Tyr1217); Phospho-Akt (Ser473). The anti-human C12orf4 antibody and all reagents were obtained from Sigma-Aldrich (St Louis, MO, USA). Antiphosphotyrosine mAb 4G10 was purchased from Upstate Biotechnology (Millipore, MA, USA). Alexa 488 conjugated anti-mouse IgG and Alexa 594 conjugated anti-rabbit IgG antibodies were purchased from Jackson ImmunoResearch laboratories (West Grove, PA, USA).

Murine bone marrow derived mast cells

To generate BMMCs, femur bones from C57BL/6 female mice (4–6 weeks old, Charles River) were isolated and progenitor cells were flushed out using a sterile protocol and cultured in Opti-MEM medium supplemented with 10% foetal calf serum (FCS), 4 mM glutamine, 100 units/ml of penicillin, 100 μg/ml of streptomycin, and 50 mM 2-mercaptoethanol together with 1 ng/ml of recombinant murine IL-3 (Biolegend). All experiments were performed in compliance with the French guidelines for experimental animal studies, and protocols were approved by the Institute of Cancer Research Ethics Committee (agreement no. B34-172-27). All reasonable efforts were made to ameliorate suffering, including anesthesia for painful procedures.

Cell culture

Culture media were obtained from Gibco (Life Technologies Ltd, Paisley, UK). Rat basophilic leukemia cell line RBL-2H3 was obtained from the ATCC (Manassas, VA, USA) and cultured in DMEM medium supplemented with 15% FCS and antibiotics. Line 293T (or HEK-T) cells were maintained in culture in DMEM medium supplemented with 10% FCS and antibiotics. The murine hybridoma 2682-1 producing anti-2,4-dinitrophenyl (DNP) IgE mAb was maintained in culture in DMEM medium supplemented with 10% FCS and antibiotics, and its culture supernatants, containing 1 μg/ml of IgE (measured by ELISA), were filtered and preserved at –20°C.

Intrabody library construction

Expression vector pEF/myc/cyto (Invitrogen, Life Technologies Ltd, Paisley, UK) was used to express the scFv library in the cytoplasm of RBL-2H3 cells. In order to clone the scFv library into this vector, VHpool and VLpool sub-libraries, which are the source of the diversity of the CDR3 loops in a previously described library [13], were assembled by PCR, cloned into the NcoI-NotI linearized vector, and transformed in *E. coli*. Library diversity was estimated to 10⁹ by counting the obtained number of clones. An aliquot corresponding to 40 times the diversity of the library was used to prepare the recombinant plasmid DNA using the Nucleobond Xtra Maxi kit (Macherey Nagel) that was subsequently used for the transfection of the RBL-2H3 cell line.

For the expression of the scFv library by retroviral infection, the scFv library was inserted between SfiI and NotI sites in

pMSCVhygSN-EGFP vector as described [14]. The estimated diversity of the library was 2×10^8 . Recloning steps after rounds 3 and 5 during selection were performed by amplification of the inserted scFv from the chromosome using pMSCV.for (cgttcgaccccgctcg) and EGFP-N.back (cgtcgcgctccagctcgaccag) primers using Phusion polymerase, followed by SfiI-NotI recloning in the same retroviral vector.

Plasmid transfection of the scFv library

RBL-2H3 cells (5×10^6) were mixed with 50 μ g of plasmid and transferred in a 4 mm electroporation cuvette (BioRad, Hercules, CA, USA). The electroporation was performed with a Gene Pulser I (BioRad) at 310 V and 960 μ F capacitance. 3×10^8 cells were electroporated for the first selection round, and 1×10^8 cells for each subsequent round (cell survival rate of 20%). For stable clone generation, cells were grown in culture medium supplemented with 1 mg/ml of geneticin (G418, Gibco).

Retroviral infection of the scFv library

Retroviral particles were produced in 293T cells using manufacturer instructions using an amphotropic envelop gene (VSV-G). Culture supernatants containing retroviral particles were collected, filtered, and used for the infection of 4×10^7 RBL-2H3 for the first selection round, and 7×10^6 cells for the recloning step after rounds 3 and 5. 48h post-infection, culture medium was replaced with fresh medium supplemented with 1mg/ml of hygromycin B (Invitrogen) as selecting agent.

Cell activation, Annexin V staining and cell sorting

Cells were incubated overnight at 37°C with anti-DNP IgE hybridoma supernatant at a final IgE concentration of 0.5 μ g/ml. Cells were washed once with RPMI, then with Tyrode buffer (10 mM HEPES pH 7.4, 130 mM NaCl, 5 mM KCl, 1 mM CaCl₂, 1 mM MgCl₂, 5.6 mM glucose, and 0.01% BSA). Cells were activated in Tyrode buffer with 100 ng/ml of DNP-KLH (keyhole limpet hemocyanin conjugated DNP, Sigma-Aldrich) at 37°C in the dark, for 45 minutes. Cells were subsequently washed in ice-cold Tyrode buffer.

For the Annexin V-APC (Becton Dickinson Biosciences, San Jose, CA, USA) staining, 100 μ l Annexin V-APC were added to 2×10^6 cells (in 500 μ l), placed 30 min on ice in the dark. The cells were then labeled with 20 μ g/ml Propidium Iodide 3 minutes prior to FACS analysis. Analysis and sorting by flow cytometry were performed using FACS Aria cell sorter (Becton Dickinson, Franklin Lakes, NJ, USA). For plasmid library selection, 2×10^8 and 10^7 cells were sorted at the first round and following rounds respectively. For retroviral selection, 4×10^7 and 10^7 cells were sorted at the first round and following rounds respectively.

High throughput sequencing

The genomic DNA of 10^6 cells was extracted and the scFv gene amplified using primers HTSVHFR3.for (nnctgtttactactgtgtgaga) and HTSVLFR4.back (nnctgttcctcccgccgaa) that hybridize to the 3' end of the FR3 of the VH and the 5' end of the FR4 of the VL respectively, and contained a 2-base index (nn in the sequence) for sequence multiplexing. This resulted in a band of 450 bp bordered by the VH and VL CDR3 regions with 18–20 bases from the flanking FR regions and a 2-base index.

Library construction was performed using the ChIPseq sample preparation kit from Illumina (IP-102–1001, San Diego, CA, USA). Briefly, 120 ng of the PCR product were repaired using T4 DNA polymerase, Klenow DNA polymerase and T4 Polynucleotide Kinase. An A was added at each 3' end followed by ligation

of Illumina's adapters. A size selection was performed on a 2% agarose gel in the 500 to 900bp range followed by an 18 cycles PCR amplification. Once purified, the library was validated using a DNA1000 chip on a BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA). Library was denatured using NaOH, hybridized on the flow cell at a concentration of 4 pM and clusterized. A 100-cycle single read sequencing was performed according to the manufacturer's instructions.

Image analyses and base calling were performed using the HiSeq Control Software (HCS 1.3.8) and RTA component (RTA 1.10.36). Using a perl script, libraries were sorted using the first 2 bases indexes (no mismatch allowed) and the next 18 to 20 bases (corresponding to the FR regions of the 2 primers) allowing one mismatch. Count of various random parts is performed using a perl script on the first 25 bases.

Sequence analysis

All analyses were done using the R statistical environment (<http://www.R-project.org>). Only the 9493 VH DNA sequences corresponding to a full CDR3 loop and present in both the naive and the final round 7 libraries were used. A two-class unpaired SAM was implemented, using the R package "samr" (<http://CRAN.R-project.org/package=samr>) [15]. In order to identify VH sequences enriched during the selection, we compared naive and round 3 (before and after the recloning step) to rounds 5, 7 and 8 libraries. If False Discovery Rate was < 0.05 , VH were considered as significantly enriched. The identified 2568 sequences were filtered by keeping those that did not contain a stop codon and whose frequency regularly increased during the selection (Naive $<$ Round 3 $<$ Round 5 $<$ Round 7), resulting in 529 VH. Finally we selected 108 VH that were enriched at least 100-fold during the selection process ($\max(\text{Round } 7, \text{Round } 8) > 100 \times \text{naive}$).

Sequences were translated and aligned using IMGT numbering [16]. Distances between sequences were calculated by giving a value of 1 if loop lengths were different and a value of (% of dissimilarity) when the two loops were of the same length (Normalized Hamming distance). Hierarchical unsupervised clustering was performed using the hclust method of R using the "complete" agglomeration method.

Measurement of β -hexosaminidase release

RBL-2H3 cells were seeded at 10^5 cells per well in 96-well culture plates. After 24 hours, adherent cells were incubated overnight with anti-DNP IgE (0.5 μ g/ml), and activated in Tyrode buffer containing 50 ng/ml of DNP-KLH for 45 minutes at 37°C, as described [17]. The release of β -hexosaminidase in the supernatant (S1) and the unreleased fraction (S2) were measured using a chromogenic substrate (p-dinitrophenyl-N-acetyl- β -D-glucosaminidase, SIGMA). The percentage of β -hexosaminidase release was calculated according to the ratio: $S1/(S1+S2) \times 100$, and expressed as a percentage of the β -hexosaminidase release of an RBL-2H3 cell line expressing an irrelevant antibody.

Measurement of TNF α secretion

RBL-2H3 cells were activated for 2 hours at 37°C as described above, and the secretion of TNF α in culture supernatants was evaluated using the Rat TNF ELISA Set (BD Biosciences, San Diego, CA, USA).

Flow cytometric analysis of Calcium mobilization and membrane FcεRI expression

For the determination of the intracellular free calcium concentration [17], 10^6 cells were stimulated with anti-DNP IgE for 2–3 hours at 37°C with gentle stirring. Cells were preloaded with 5 μM Fluo-3AM (Molecular Probes, Life Technologies Ltd) in the presence of 0.2% Pluronic F-127 (Molecular Probes) for 30 minutes. Cells were activated by the addition of DNP-KLH at a final concentration of 200 ng/ml and the intracellular free calcium concentration was monitored with a FC500 flow cytometer (Beckman Coulter, Inc. Brea, CA, USA).

For the evaluation of surface expression of FcεRI, cells were incubated for 2 h at 37°C with anti-DNP IgE. The membrane-bound IgE was detected using biotinylated anti-mouse Ig (BioLegend, San Diego, CA, USA) followed by FITC-conjugated streptavidin (GE Healthcare, Buckinghamshire, UK).

Production and purification of antibody fragments

Antibody fragments were expressed in *E. coli* after cloning in pET23NN vector as described [13]. The resulting antibody fragment is tagged with a c-Myc and a 6xHis tag at its C-terminus. For pull-down experiments, antibody fragments were purified from bacterial cytoplasmic extracts using magnetic nickel beads (Ademtech, Pessac, France).

The production of bivalent antibodies used in immunofluorescence experiments was achieved by cloning antibody fragments between BglII and EcoRI site of vector ps1119 [18], allowing the production of N-terminal fusions to a mouse Fc of the IgG1 isotype. 293T cells were transiently transfected using JetPEI (Polyplus, NY, USA) and grown for 6 days. Culture supernatants enriched in antibody-Fc fusion were harvested, filtered on 0.2 μm and stored at –80°C.

Target pull-down

Cells were incubated overnight at 37°C with anti-DNP IgE and they were activated with 50 ng/ml of DNP-KLH for 3–10 minutes at 37°C as described above. After washing with cold PBS containing phosphatase inhibitors (100 mM sodium fluoride, 5 mM sodium orthovanadate), cells were lysed for 15 minutes on ice with lysis buffer (PBS supplemented with 0.5% sodium deoxycholate, 1% NP-40, 0.1% SDS). Cell lysates were clarified by centrifugation for 15 minutes at 4°C at 16,000 g. The total protein content of the soluble fraction was quantified using the BCA assay kit (Interchim, Montluçon, France).

For pull-down experiments, 3 mg of protein lysates were incubated with 20 μl of magnetic nickel beads loaded with 20 μg of purified antibody fragment (see above) for 2 h at 4°C. Beads were washed 3 times in lysis buffer supplemented with 10 mM imidazole. Elution was performed by the addition of 500 mM imidazole to the beads. Eluates were analyzed by SDS-PAGE.

SDS-PAGE and MS/MS analysis

Captured proteins were separated by SDS-PAGE (10%) and detected with Coomassie-brilliant blue staining. Bands of interest were cut off from the gel. Gel pieces were subjected to in-gel digestion with trypsin (Promega, Fitchburg, WI, USA) as described [19]. Desalted peptides were resolved on an Ultimate 3000 nano-LC System (Dionex, Thermo Fischer Scientific) equipped with a PepMap 100 C18 column (3 μm particles, 10 nm pore size, 75 mm id ×15 cm). For MALDI MS/MS analysis, column effluent was mixed in a 1/3 ratio with MALDI matrix (2 mg/ml acyano-4-hydroxycinnamic acid (LaserBio Labs, CA, USA) in 0.1% TFA, 70% acetonitrile) and spotted on an Opti-TOF 384-well

Insert 123×81mm plate. MALDI plates were analyzed using the 4800 Plus MALDI TOF/TOF Proteomics Analyzer mass spectrometer (AB Sciex, MA, USA) in positive reflector ion mode. Each MS spectrum is the result of 1500 averaged laser shots. In MS/MS mode, fragmentation of the 12 most intense selected precursors was performed at collision energy of 2 kV, each MS/MS spectrum is the result of 3000 laser shots. Protein identifications were performed in Uniprot/Swiss-Prot2012_01 database by ProteinPilot Software V 2.0.1 (AB Sciex) using the Paragon algorithm. This software calculates a confidence percentage that reflects the probability that the hit is a false positive, meaning that at 99% confidence level (unused score >2), there is a false positive identification chance of about 1%. After database searching, only proteins identified with an unused score ≥2 and peptides identified with a confidence score ≥95 were retained.

Western blot

Following SDS-PAGE electrophoresis, proteins were transferred on a 0.2 μm nitrocellulose membrane. Before each hybridization, the membrane was blocked in 5% BSA in TBS-T buffer (10 mM Tris pH 7.4, 150 mM NaCl, and 0.1% Tween) for 1 h at room temperature. Proteins were revealed using primary and secondary antibodies coupled to peroxidase according to manufacturer's recommendation. For the signal detection, the ECL-Plus chemiluminescent substrate (PerkinElmer, Waltham, MA, USA) and a camera (G-Box, Syngene, Cambridge, UK) were used. Signal intensities were quantified using the manufacturer-supplied software.

Immunofluorescence

The cells were seeded on glass slides in Labtek chambers (Nunc, Thermo Fischer Scientific). All stages of these experiments were performed at room temperature. After two washes in PBS, cells were fixed with 3.7% paraformaldehyde (Sigma Aldrich) for 10 minutes, and then permeabilized with PBS containing 0.1% Triton X-100 and 1.5% FCS for 10 minutes. Cells were incubated with the primary and the fluorescent antibodies for 1 to 2 h at room temperature. 5H4-VH-Fc fusion protein and a commercial rabbit anti-C12orf4 antibody were detected by Alexa 488 conjugated anti-mouse IgG and Alexa 594 conjugated anti-rabbit IgG antibodies, respectively. After washing, the slides were mounted in Mowiol, then visualized and captured using a Zeiss LSM 510 Meta Confocal Microscope (Oberkochen, Germany).

shRNA

The two shRNA (sh1: cattctaattctctcgaaa; sh2: agaattgattggc-gaaaga) were cloned into vector pSIREN (Clontech, Mountain View, CA, USA). Retroviral supernatants were produced as described above. Infected RBL-2H3 cells were selected by addition of 2.5 μg/ml of puromycin (HyClone, Fischer Thermo Scientific) to the culture medium two days after retroviral infection.

RT-qPCR

Total RNA was extracted and purified from 10^5 – 10^6 cells using a Qiagen kit. 1 μg of RNA was reverse transcribed using 100 ng of random primer and the M-MuLV Reverse Transcriptase (Invitrogen). qPCR was performed on the cDNA using 2 pairs of primers (ccaaagcgtatgctgagaca/ctgcatcaccctttccatt and ctggaaccaaagt-gagga/cgagcagtgatgttctctga) and SYBR Green I Master mix on a Light Cycler 480 (Roche, Basel, Switzerland). The data were analyzed using the software supplied by the same manufacturer.

Results

We performed the phenotypic selection in the rat basophilic leukemia cell line RBL-2H3, used extensively to study FcεRI and the biochemical pathways for secretion in mast cells. First, we subcloned a single chain antibody fragment (scFv) library optimized for intracellular expression [13,20] in plasmid and retroviral eukaryotic expression vectors, both designed for cytoplasmic expression of intrabodies, and two libraries of a diversity of 10^9 and 2×10^8 were generated respectively. The recombinant vector pools were subsequently used to transfect the RBL-2H3 cell line in order to generate two distinct populations of 5×10^7 transformed cells. Previous studies reported an Annexin-V binding assay as a powerful method to monitor mast cell degranulation for functional analysis [21]. The IgE-dependent stimulation of the cells leads to the exposure of exocytosing granules and phosphatidylserines that can be monitored, in proportion to the extent of allergic mediator release, by the binding of exogenously added Annexin-V. We used this method to isolate by flow cytometry cell sorting (FACS) the population of intrabody-containing RBL-2H3 cells that displayed an impaired degranulation.

Both transfection approaches have advantages and drawbacks. Using plasmid expression allows a higher expression level with the inconvenience of several different intrabodies expressed in the same cell. We reasoned that in the case of a dominant phenotype this should not preclude the selection of inhibitory molecules. On the opposite, using retroviral transfection ensured that only few different intrabody genes were present in each cell, presumably allowing a more efficient selection. However, in the latter case the explored diversity was limited to the number of transfected cells (5×10^7), whereas in the former case the diversity was the product of the cell number by the number of plasmids per cell (2000 copies) and was thus only limited to the size of the initial library (10^9). Thus, we opted for using both approaches because this would allow not only to explore the limiting factors of each system but also to compare the results obtained from two independent selections.

Plasmid library selection

In the case of the plasmid library, seven rounds of enrichment were performed based on a selection method comprising the following steps: a) FcεRI-mediated cell stimulation followed by Annexin V-APC staining; b) FACS sorting of 10% of the population corresponding to the less fluorescent cells; c) extraction of the plasmid DNA pool; d) amplification of the scFv genes and their subsequent cloning in the same expression vector (Fig. 1a). This procedure allowed the enrichment for cells containing intrabodies able to block cell degranulation as shown by the decrease of the Annexin V staining of stimulated versus unstimulated cells (Fig. 1b, Fig. S1a). In addition, because each round of selection included a recloning step of scFv genes, this procedure ensured that the observed phenotype was specifically due to the transfected intrabody genes and not to a cell drift. Another issue with the selection was the possible anti-apoptotic effect of monomeric IgE used for FcεRI stimulation [22,23]. However, since we have incubated both the unstimulated and the stimulated cells with IgE before the addition of the antigen DNP, if present this effect should be identical in both populations and should not have influenced the selection.

To identify individual intrabody sequences responsible for the inhibitory phenotype, we generated stable RBL-2H3 clones from the pool of plasmids obtained from the seventh round of selection. Using qPCR, we estimated that during the selection, each cell contained an average of 2,000 plasmids, thus we reasoned that

most of the cells should contain non-inhibitory passenger intrabodies. One hundred and twenty-six stable clones were tested in a degranulation assay based on the measurement of the FcεRI-mediated release of the enzyme β-hexosaminidase. As shown in Fig. 2a, the β-hexosaminidase release values are distributed following a bimodal distribution with a major peak at 110% and a smaller one at 65%. The former peak contains cells expressing non-inhibitory intrabodies whereas the latter corresponds to the low-degranulating clones that express inhibitory intrabodies and which represent about 20% of the clones. The analysis of the intrabody sequences expressed in 36 clones revealed a high diversity with 1–2 different sequences expressed in each clone (Fig. S2).

Retroviral library selection

In the case of the retroviral library, the screen was performed in the same manner. Since retroviral infection directly generated stable clones, extraction of the intrabody genes and recloning between each round was not necessary. We however introduced a recloning step after the third and the fifth round of selection. This ensured that the retrovirus-induced phenotype was associated with the expressed intrabody sequence and was not due to a cell drift or a particular genomic insertion site. Inhibition of the phenotype was clear after seven rounds of selection as the shift in Annexin V of FcεRI-stimulated cells was totally abolished at round 7 (Fig. 1b & Fig. S1b). An aliquot of cells from selection round 7 was seeded at limiting dilution and 48 isolated clones were analyzed in triplicate by measuring the release of β-hexosaminidase (Fig. 2b). Contrary to the plasmid selection, the distribution of the β-hexosaminidase release values was monomodal and significantly inhibited by 54% on average when compared to 11 mock-transfected clones. This showed that the retroviral library selection was more powerful than the plasmid one, presumably because of the presence of fewer intrabodies per cell, which reduced the co-selection of passenger intrabodies to a minimal.

We used high throughput sequencing for the analysis of the intrabody diversity evolution in the cell population during the course of selection and for the identification of the best inhibitory intrabodies. This avoided a biased analysis of a limited number of cellular clones that may bear peculiar properties. scFv sequences expressed in the initial pool of RBL-2H3 infected cells before the first selection (naive library), and from cells from each of rounds 3, 5, 7 and 8 were analyzed. Sequencing of 10^6 clones from the naive library revealed that 80% of the sequences all had H3 or L3 without frameshift or stop codon. The translation of the nucleotide sequences generated 250,000 unique VH amino acid sequences and 350,000 VL sequences. This reflects the actual diversity of the library contained in 10^6 cells but do not take into account the additional diversity due to the random pairing of VH and VL chains (Fig. S3).

Since the VH domain is known to be the most important determinant of antibody affinity and specificity [24], we only analyzed the CDR3 sequences that correspond to the variable part of the VH domains in our library design [13]. By setting the False Discovery Rate to 0.05, 2568 VH DNA sequences appeared to be enriched and none significantly depleted during the course of the selection using the SAMSeq software [15]. The fact that no sequence declined significantly during the selection shows that none of the intrabodies was toxic enough to the cell to induce an early apoptotic phenotype and to be selected. Enriched sequences were translated and we retained the 108 VH sequences that continuously increased during the course of the selection. These 108 VH represented 40% of the sequences present in the final selected library and were grouped in 69 families using unsuper-

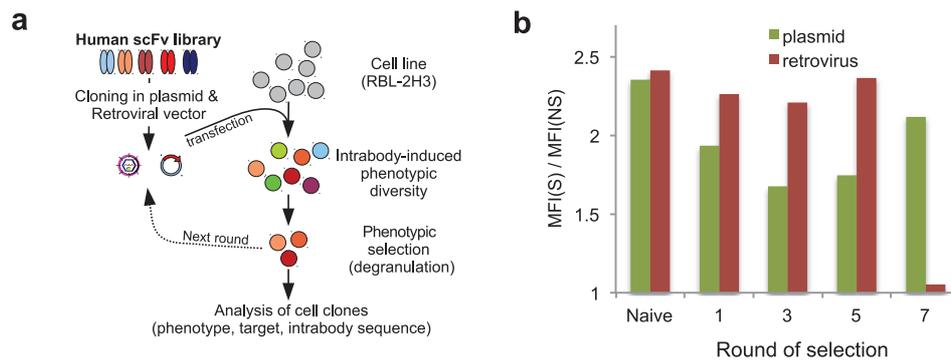


Figure 1. Selection of intrabodies that inhibit mast cell degranulation. a) Schematic view of the selection method. The scFv/intrabody library previously described [13] was cloned in plasmid and retroviral vectors and used to transfect the RBL-2H3 cell line in order to induce a phenotypic diversity in a collection of cells. Clones displaying the desired phenotype, measured by inhibition of degranulation, were selected and the couple constituted by the inhibitory intrabody and its target antigen was identified and characterized. b) Annexin-V staining of cell populations from the library selection rounds is illustrated as the ratio of the geometric mean (MF) of the FcεRI-stimulated (S) to the unstimulated (NS) cells (Fig. S1). doi:10.1371/journal.pone.0104998.g001

vised hierarchical clustering and a cutoff of 60% identity (Fig. S4). Thirty-nine families contained a unique sequence and were not further analyzed. The remaining 69 VH clustered in 17 groups, and we retained the 10 VH families present at least at 0.1% (0.1–25%) in the final enriched library. Between the naive library and the final selected library, each of these 10 families was enriched about 150 fold (90–215), however some individual sequences were enriched more than 500-fold. In addition, whereas the frequency of most of the clones present in the initial library decreased at an exponential rate during the course of the selection, the 10 identified families were strongly enriched (Fig. 2c & Fig. S5).

Characterization of individual clones

Next, 136 unique VH sequences obtained from 178 clones of the plasmid library (stable clones and randomly picked sequences from the last round of selection) were compared with the 108 VH sequences enriched during the retroviral selection. Only one VH sequence corresponding to clone 5H4 was found common to both selections. Moreover, 5H4 family (R_8) is part of the 10 best

families selected from the retroviral sequence analysis and the 5H4-VH sequence was enriched 170 fold during the retroviral selection. Plasmid-derived stable RBL-2H3 clone 5H4 did not show any Annexin-V staining following FcεRI stimulation (Fig. 3a) and displayed strong defects in β-hexosaminidase release, calcium flux and TNFα secretion (Fig. 3b, 3c, 3d). Analysis of 5H4 sequence revealed that the gene encoding scFv 5H4 was truncated at its C-terminus because of the presence of a stop codon in the first codon of the light chain CDR3. Since the original scFv library contained only variable CDR3 loops, we reasoned that the inhibitory phenotype of 5H4 was carried by its VH portion. Indeed, isolated human VH domains have already been shown to be efficient intrabodies, particularly when the VH sequence belongs to the human VH3 family as it is the case here [25,26]. In fact, the analysis of the FcεRI-mediated degranulation of the retroviral clone R_8 confirmed the inhibitory potential of the intrabody 5H4-VH (Fig. 3e).

We also evaluated the inhibitory phenotype of the 9 other families of intrabodies selected during the retroviral selection. The most frequent sequence of each family (Fig. S6) was cloned and the

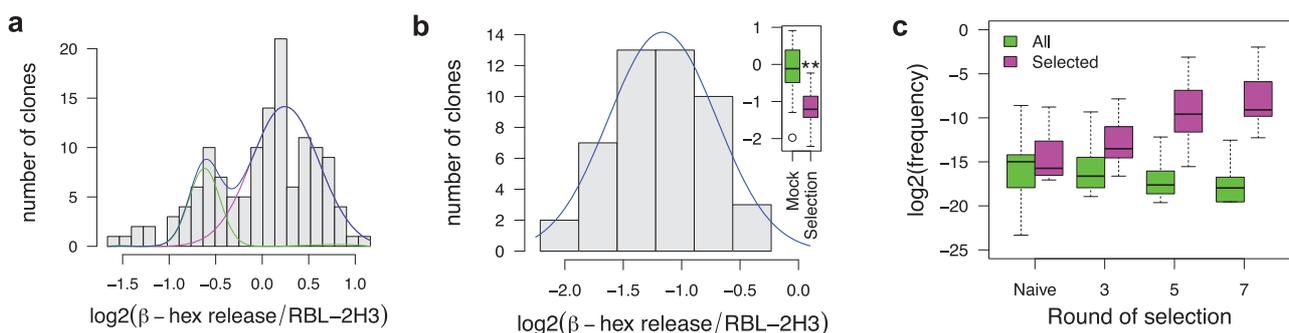


Figure 2. Analysis of selected clones. a) Distribution of the β-hexosaminidase release measured on 126 stable clones isolated from the last round of the plasmid selection. The distribution profile does not fit a normal distribution ($p = 0.027$ using Jarque-Bera Normality Test) and is skewed to the left ($p = 0.009$ using Agostino's skewness test [53]). The blue curve is the sum of the two normal distributions plotted in magenta and green and was fitted to the distribution. b) Distribution of the β-hexosaminidase release measured on 48 retroviral clones. The distribution is normal ($p = 0.92$) and the blue curve is the best normal distribution fitted to the data. Inset: boxplot of the β-hexosaminidase release of 48 retroviral clones compared to 11 mock-transfected RBL-3H2 clones. **: $p < 0.01$ (Student t-test). Boxplot whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range. c) VH sequence evolution during retroviral selection followed by high throughput sequencing. All: frequency of the 6789 VH sequences present in the four sequenced pools. Selected: enrichment of the 125 DNA sequences (62 different CDR3 in amino acid) forming the 10 selected families (Fig. S4 & S5). Boxplot whiskers extend to the most extreme data point. doi:10.1371/journal.pone.0104998.g002

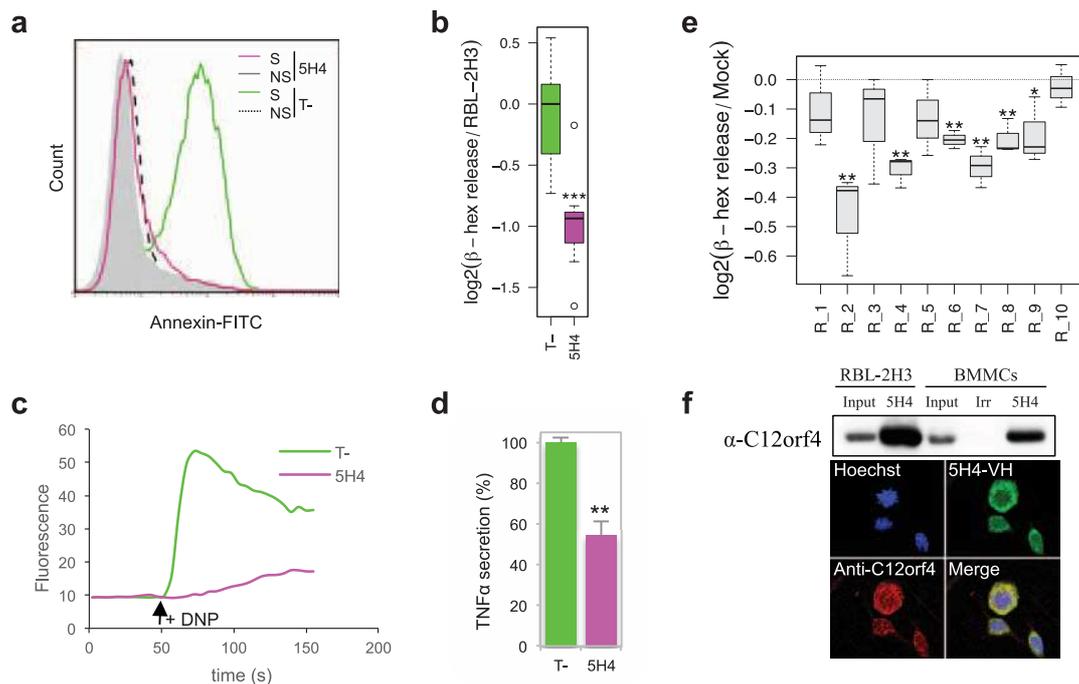


Figure 3. Anti-C12orf4 intrabody inhibits mast cell degranulation. Analysis of stable clone 5H4: a) measurement of Annexin-V staining; b) β -hexosaminidase release; c) calcium flux and d) $\text{TNF}\alpha$ secretion. T-: Irrelevant intrabody. S: IgE/DNP stimulated. NS: unstimulated. Boxplot whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range. e) Measure of β -hexosaminidase release by retroviral infected populations. Clone R_8 is identical to the intrabody expressed by the plasmid clone 5H4. Sequences of the clones are given in Fig. S6. Boxplot whiskers extend to the most extreme data point. f) Specific binding of 5H4-VH to C12orf4. Top panel: pull-down assay using 5H4-VH as capture agent and a commercial anti-C12orf4 polyclonal serum to reveal the protein. Irr: Irrelevant VH fragment, differing from 5H4 VH only by its CDR3 sequence. Low panel: subcellular localization of C12orf4 analyzed by confocal laser microscopy after double staining. Top left: Hoechst; top right: 5H4-VH-Fc fusion; bottom left: anti-C12orf4 commercial antibody; bottom right: merge. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$ (Student t-test). doi:10.1371/journal.pone.0104998.g003

cellular phenotype of retrovirus-infected mixed cell populations was analyzed. As shown in Fig. 3e, 6 out of 10 intrabodies induced a significant inhibition of degranulation following $\text{Fc}\epsilon\text{RI}$ stimulation. These results confirmed that our statistical analysis of the data generated by high throughput sequencing successfully led to the identification of inhibitory intrabody sequences.

Target identification

The strength of the IBPhES approach lies in the fact that intrabodies are antibody molecules bearing high affinity and specificity which can be used for the identification of their target. In order to identify the cellular target of intrabody 5H4, the gene encoding 5H4-VH was expressed in *E. coli* cytoplasm, and the purified antibody fragment was used to capture its target from RBL-2H3 extracts. The captured proteins were analyzed by mass spectrometry using an irrelevant VH fragment as a control. A unique protein called RGD1311164, the homolog of human protein C12orf4, was identified in three independent experiments with the best score. For an easier reading we will refer to RGD1311164 as C12orf4. The specific binding of 5H4-VH to C12orf4 from rat and mouse origin was confirmed in pull-down experiments performed respectively on RBL-2H3 cell line and murine bone marrow-derived mast cells using a commercial polyclonal serum (Fig. 3f & Fig. S7a). Because 5H4 VH differs from the irrelevant VH fragment only by its CDR3 sequence (10 amino acids), its specificity for C12orf4 is driven by its CDR3 sequence and cannot be due to the exposed hydrophobic VH-VL interface. Immunoprecipitation experiments as well as immuno-

fluorescence analysis using confocal microscopy with either 5H4-VH-Fc or a commercial anti-C12orf4 antibody showed that C12orf4 is a cytoplasmic protein, and that its subcellular localization and expression level does not change upon $\text{Fc}\epsilon\text{RI}$ stimulation (Fig. 3f & Fig. S7b).

Characterization of RGD1311164/C12orf4

We further characterized the role of C12orf4 in $\text{Fc}\epsilon\text{RI}$ -mediated mast cell responses. For this purpose, the short hairpin RNA (shRNA) approach was used for down regulation and modulation of C12orf4 expression in RBL-3H2 cells. As shown in Fig. S8a, the two shRNA decreased the mRNA level by 60–80% and the protein level by 55–70%. The analysis of the degranulation of shRNA transfected cells following $\text{Fc}\epsilon\text{RI}$ stimulation showed a decrease in β -hexosaminidase release and $\text{TNF}\alpha$ secretion that correlated with an inhibited calcium flux (Fig. S8b). These results confirmed using an independent approach that inhibition of C12orf4 leads to a defect in mast cell activation.

In mast cells and basophils, the engagement of $\text{Fc}\epsilon\text{RI}$ initiates the activation of the Src kinases Lyn and Fyn and the Syk kinase, which allows signal propagation through the phosphorylation and activation of several downstream proteins [27]. Lyn phosphorylates Syk that coordinates further signals such as the activation of PLC- γ and calcium mobilization. Fyn initiates a complementary signaling pathway through the adapter Gab2 that is essential for the activation of PI3K and cell degranulation (Fig. 4b). We investigated the impact of targeting C12orf4 on $\text{Fc}\epsilon\text{RI}$ -mediated signaling events. Protein extracts of non-stimulated and $\text{Fc}\epsilon\text{RI}$ -

stimulated stable clone 5H4 were analyzed by western blot using anti-phosphotyrosine monoclonal antibody 4G10. A defect in the FcεRI-mediated tyrosine phosphorylation of proteins migrating at 40–42, 55–60 and 72 kDa was observed. Using antibodies specific to key proteins involved in the signaling pathways, we confirmed a decreased phosphorylation of Src and Syk tyrosine kinases as well as the downstream proteins MAPKs and Akt, demonstrating an impairment of both Lyn- and Fyn-dependent signals (Fig. 4a). These results are consistent with the defect in the degranulation events (Fig. 3b, 3d). The analysis of the protein extracts of cells transfected with a C12orf4-targeting shRNA showed a milder overall effect, with a selective modulation of the Fyn pathway as shown by the decreased phosphorylation of downstream proteins Gab2 and Akt (Fig. S8c). This may be due to the moderate shRNA-mediated inhibition of C12orf4 expression (about 50%). Taken together, our results suggest that C12orf4 plays a role in the early signaling events following FcεRI-induced cell activation.

Discussion

In this study, we report the identification of protein RGD1311164/C12orf4 as a new player in FcεRI-induced cell degranulation. Database interrogation revealed that C12orf4 is a protein of unknown function. In human, C12orf4 gene is localized on chromosome 12p13.3. Sequence analysis showed that C12orf4 is widely conserved from nematodes to humans, with for instance 94% amino acid identity between rat and human protein sequences. This suggests a common role in all these organisms, but none of the proteins of this family have a defined function or homology to a domain of known function. Genini and co-workers [28] suggested a possible link between C12orf4 gene and the autosomal recessive disease known as arthrogryposis multiplex congenita (AMC), one of the most common congenital defects observed in pigs and in other mammals.

Our data suggest that C12orf4 is a cytoplasmic protein and that its cellular localization is not affected following FcεRI stimulation. RBL-2H3 stable clones expressing anti-C12orf4 5H4 inhibitory intrabody, as well as cells expressing a shRNA against C12orf4,

showed an inhibition of mast cell activation and a decrease in allergic mediator release. These cellular responses correlate with defects in the early signaling events following FcεRI stimulation, as shown by a decrease in the phosphorylation of key upstream proteins such as Src kinases. Because the level of FcεRI expression at the membrane of transfected cells was not affected, C12orf4 is presumably acting upstream of Src kinases, but is not implicated in receptor recycling and degradation. We analyzed 5H4-pull down experiments with antibodies specific to FcεRI, Lyn, Syk and with the anti-phosphotyrosine monoclonal antibody 4G10, and none of the corresponding targets were detected. This is not uncommon since some signaling molecules interact weakly with their partners and it is sometimes difficult to demonstrate co-immunoprecipitation of endogenous proteins. Another hypothesis is that 5H4 may inhibit the interaction of C12orf4 with its cellular partners, precluding detection of interacting proteins in pull-down experiments.

The IBPheS method described here allowed us to identify several other intrabodies that affect the degranulation of RBL-2H3 cells. Presumably, some of these intrabodies recognize known targets implicated in the degranulation process. However, the diversity of the used intrabody repertoire is somewhat limited because of the difficulty to manipulate more than 10^8 cells, and thus the approach cannot allow a complete enumeration of all the players implicated in the studied phenotype. This however demonstrates that the selection of intrabodies as phenotype-modulators can be accomplished in cell-based assays using unbiased combinatorial libraries (not preselected by phage panning) and without any previous knowledge of a target. On this line, two recent studies describe the use of large and naïve intracellular combinatorial libraries for phenotypic modulation. The first study reports the selection of secreted integrin-binding agonist antibodies that convert stem cells to dendritic cells [29]. The second approach describes the use of lentiviral libraries of intrabodies for the selection of scFv fragments that block rhinovirus-induced cell death [30]. In this case, the studied phenotype is highly selectable because based on cell survival. This explains the very impressive enrichment factor of more than one

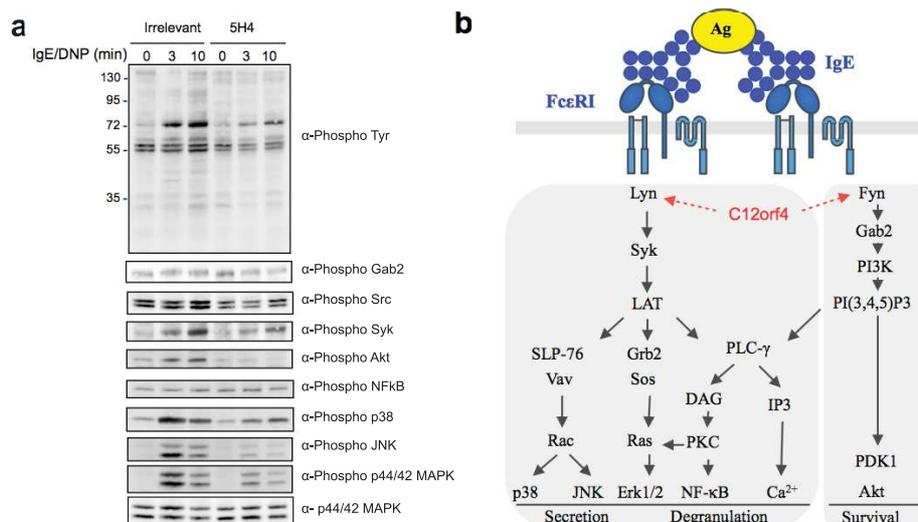


Figure 4. C12orf4 is implicated in the early events of the degranulation pathway. a) Western blot analysis of the FcεRI-mediated phosphorylation of major proteins implicated in mast cell activation. 5H4 expressing RBL-2H3 cells are compared to cells expressing an irrelevant antibody fragment. Cells are either non-activated or activated with IgE/DNP for 3 and 10 minutes as indicated. b) Schematic view of mast cell signaling pathways, kindly provided by Dr. Marc Daëron. doi:10.1371/journal.pone.0104998.g004

million-fold since only the cells containing protective intrabodies survived viral infection. In the present work, we show that despite a weaker FACS-based phenotype, the selection is efficient enough to isolate inhibitory intrabodies. This makes the approach applicable to a larger number of cellular phenotypes. Indeed, when coupled with high throughput sequencing, the method is able to identify very rare clones. In fact, the sequence analysis revealed that intrabodies that were present at a frequency of 10^{-7} in the naive library were significantly enriched and identified in the 10 families of retroviral clones (Fig. S5). Thus, as in the case of phage-display selection, the use of high throughput sequencing allowed the identification of clones even expressed well below the background and impossible to select using a direct cloning strategy [31].

If we compare IBPheS to classical genetics, the intrabody plays the same role than mutations. However, because the intrabody does not directly modify its target but modulates its function, it is also analogous to a chemical drug. Thus the system described here has the advantages of both approaches, that is the power and flexibility of genetics coupled with the clinical applications of pharmacology. Intrabodies have demonstrated they are able to recapitulate all the antibody properties within the cell: enzyme inhibition [32]; breaking protein-protein [33] and protein-DNA interactions [34]; re-activating mutant enzymes [35]; targeting specific protein conformations [36,37], domains [17], and post-translational modifications [38]; and inducing protein degradation [39,40]. In addition, by targeting them to specific cell compartments intrabodies can re- or de-localize their target [41] and block secretion [42]. Intrabodies are thus able to mimic the whole spectrum of mutations that can be obtained in classical genetics. However, in addition to proteins, antibodies are also able to recognize small chemicals, usually referred as haptens, glycans and lipids [43,44]. As such they represent a powerful tool not only to interrogate the proteome diversity but also secondary messengers and metabolism in cells. As such the IBPheS method must be seen as a complement to other methods such as genome-wide shRNA screens [45–47], but with specific advantages associated with the direct targeting of effector proteins instead of mRNA.

Intrabodies have also proven their potential to be used in clinics [48]. Delivery of an anti-erbB2 intrabody using an adenovirus vector has been described in a phase I clinical trial with minimal toxicity [49]. The main current limitation is intrabody gene delivery but this will improve with advances in gene transfer technology. An alternative approach still not demonstrated in the case of intrabodies could be the use of internalizing peptides or liposomal vehicles to directly deliver proteins within tissues in pre-clinical models and in patients [50]. However, since this has already been described in cell cultures [51,52], this route of intrabody delivery could be a viable solution. As an easier and more generally applicable solution, we reported the use of an antibody displacement assay to convert an intrabody directed against the tyrosine kinase Syk into chemical drugs [7,8]. The isolated molecules recapitulated the intrabody effects in cell cultures and were able to block the anaphylactic shock when administrated orally in animal models [7].

In conclusion, the IBPheS method aims to be an integrated approach for the concomitant identification of a protein target and an intrabody as a lead inhibitor. As such, and compared to other large-scale approaches, this represents a straightforward path to the discovery of potential therapeutic molecules.

Supporting Information

Figure S1 FACS analysis of the IgE/DNP stimulated (S: green) and non-stimulated (NS: black) RBL-2H3 cells transformed with plasmid (a) or retroviral (b) libraries. Naive: unselected library; Round n: enriched library after n rounds of selection. The X axis represents Annexin V labelling, and the Y axis the Forward Scatter (FCS). (PDF)

Figure S2 Individual clone phenotypes from plasmid library selection. a) 133 stable clones were tested for β -hexosaminidase release. Red dots represent the clones selected for sequencing. The clone 5H4, characterized in Fig. 3 of the manuscript, is marked. b) 36 clones were sequenced. Clones are sorted from the least to the most degranulating clone in the initial screen in (a). Nb seq: number of intrabody sequences retrieved by PCR. “> = 2”: the clone contains more than 1 sequence but was not analyzed further to determine the exact number of inserted intrabody. H3: VH CDR3 sequence. L3: VL CDR3 sequence. *: stop codon. x: unread because of poor sequencing quality. c) Clones for which a sequence was determined, were re-tested for β -hexosaminidase release (between 2 and 6 replicates). *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$ (Student t-test). Clones 6E8, 8D12 and 5F5 were used as negative controls. (PDF)

Figure S3 Intrabody library diversity. Sequence analysis of DNA extracted from one million clones infected with the indicated retroviral libraries. Nb of reads: number of reads for each library; Nb of seq (dna): Number of different full length CDR3 DNA sequence; Nb of seq (dna no stop): Number of different full length CDR3 DNA sequence without stop codon or frameshift; Nb of seq (aa): Number of different CDR3 protein sequences obtained without stop codon or frameshift. a) VH CDR3. b) VL CDR3. Round3a and Round3 are the same pool but sequenced before and after recloning respectively (see Materials and Methods). (PDF)

Figure S4 Clustering of VH sequences identified in retroviral library selection. Sequence analysis of the 108 VH enriched during retroviral selection. Sequences are aligned according to IMGT numbering scheme. The two numbers at the left of each sequence are the frequency in the selected library (round 7) and the enrichment factor between naive and final library, respectively. The 10 retained families after clustering are indicated. (PDF)

Figure S5 Population evolution of the clones from the 10 selected families. For each family in Fig. S4, the evolution of the frequency of all the clones is plotted. The sequence above the plots is the VH CDR3 sequence of the most abundant clone used in the validation study in Fig. 3e. Since clones were considered as different when their DNA sequences were different, the number of clones in each family does not necessary match the number in Fig. S4 that compared translated CDR3 sequences. (PDF)

Figure S6 Sequences of the most abundant retroviral clones from the 10 selected families. R_7 has a stop codon in the VL CDR3 loop and is thus truncated and expressed without the C-terminal eGFP tag. R_8 has the same VH than plasmid clone 5H4 and has been cloned as a single VH domain in retroviral vector. (PDF)

Figure S7 Analysis of FcεRI-induced C12orf4 expression and subcellular localization. a) Pull-down assays were performed on total lysates of non-stimulated and FcεRI-stimulated RBL-2H3 cells, using 5H4-VH and an irrelevant VH fragment. The presence of C12orf4 in protein extracts and pull-down fractions (PD) was detected using a rabbit anti-C12orf4 polyclonal serum. b) Analysis of subcellular localization of C12orf4 following FcεRI-stimulation. RBL-2H3 cells were either non stimulated (top panels) or stimulated for 3 minutes (middle panels) and 10 minutes (bottom panels) with IgE/DNP as described in methods and stained with a commercial rabbit anti-C12orf4 serum followed by an anti-rabbit IgG Alexa 599 labeled secondary antibody (left panels). Nuclei were stained with Hoechst (right panels). (PDF)

Figure S8 shRNA-induced down-regulation of C12orf4 expression. a) Two shRNA against rat C12orf4 (sh1 and sh2) were cloned in a retroviral vector and transduced in RBL-2H3 cells. Analysis of C12orf4 expression by qPCR and western blot were performed 10 and 15 days post-infection. b) Analysis of β-hexosaminidase release (left), calcium flux (middle), and TNFα

secretion (right) were performed with cell populations 5 days post-infection. c) Western blot analysis of the FcεRI-mediated phosphorylation of major proteins implicated in mast cell activation. Cell populations transfected with sh1 C12orf4 (5 days post-infection) are compared with a control shRNA (shLUC), either non activated or activated with IgE/DNP for 3 and 10 minutes. **: p<0.01; ***: p<0.001 (t-test). (PDF)

Acknowledgments

We are grateful to Dr. M. Daéron (Institut Pasteur, Paris, France) for critically reading the manuscript and for useful comments. We thank Dr. C. Duperrey and Dr. E. Jublanc for their expertise in using flow cytometry and confocal microscope facilities at Montpellier Rio Imaging platform and Dr. P. O. Harmand for technical assistance.

Author Contributions

Conceived and designed the experiments: PD PM CH RJ. Performed the experiments: PD PM EM LG VP NB CL HP JPD CM. Analyzed the data: PD PM CH RJ. Contributed to the writing of the manuscript: PD PM.

References

- Galli SJ, Kalesnikoff J, Grimaldeston MA, Piliponsky AM, Williams CMM, et al. (2005) Mast cells as “tunable” effector and immunoregulatory cells: recent advances. *Annu Rev Immunol* 23: 749–786. doi:10.1146/annurev.immunol.21.120601.141025.
- Bounab Y, Getahun A, Cambier JC, Daéron M (2013) Phosphatase regulation of immunoreceptor signaling in T cells, B cells and mast cells. *Curr Opin Immunol* 25: 313–320. doi:10.1016/j.coi.2013.04.001.
- Riccaboni M, Bianchi I, Petrillo P (2010) Spleen tyrosine kinases: biology, therapeutic targets and drugs. *Drug Discov Today* 15: 517–530. doi:10.1016/j.drudis.2010.05.001.
- D’Cruz OJ, Uckun FM (2012) Targeting Spleen Tyrosine Kinase (Syk) for Treatment of Human Disease. *J Pharm Drug Deliv Res* 01: doi:10.4172/2325-9604.1000107. doi:10.4172/2325-9604.1000107.
- Meltzer EO, Berkowitz RB, Grossbard EB (2005) An intranasal Syk-kinase inhibitor (R112) improves the symptoms of seasonal allergic rhinitis in a park environment. *J Allergy Clin Immunol* 115: 791–796. doi:10.1016/j.jaci.2005.01.040.
- Norman P (2009) A novel Syk kinase inhibitor suitable for inhalation: R-343(?)–WO-2009031011. *Expert Opin Ther Pat* 19: 1469–1472. doi:10.1517/13543770903059281.
- Mazuc E, Villoutreix BO, Malbec O, Roumier T, Fleury S, et al. (2008) A novel druglike spleen tyrosine kinase binder prevents anaphylactic shock when administered orally. *J Allergy Clin Immunol* 122: 188–194, 194.e1–3. doi:10.1016/j.jaci.2008.04.026.
- Villoutreix BO, Laconde G, Lagorce D, Martineau P, Miteva MA, et al. (2011) Tyrosine kinase syk non-enzymatic inhibitors and potential anti-allergic drug-like compounds discovered by virtual and in vitro screening. *PLoS One* 6: e21117. doi:10.1371/journal.pone.0021117.
- Dauvillier S, Merida P, Visintin M, Cattaneo A, Bonnerot C, et al. (2002) Intracellular single-chain variable fragments directed to the Src homology 2 domains of Syk partially inhibit Fc epsilon RI signaling in the RBL-2H3 cell line. *J Immunol* 169: 2274–2283.
- Gargano N, Cattaneo A (1997) Rescue of a neutralizing anti-viral antibody fragment from an intracellular polyclonal repertoire expressed in mammalian cells. *FEBS Lett* 414: 537–540.
- Gargano N, Cattaneo A (1997) Inhibition of murine leukaemia virus retrotranscription by the intracellular expression of a phage-derived anti-reverse transcriptase antibody fragment. *J Gen Virol* 78 (Pt 10): 2591–2599.
- Cattaneo A, Biocca S (1999) The selection of intracellular antibodies. *Trends Biotechnol* 17: 115–121.
- Philibert P, Stoessel A, Wang W, Sibley A-P, Bec N, et al. (2007) A focused antibody library for selecting scFvs expressed at high levels in the cytoplasm. *BMC Biotechnol* 7: 81. doi:10.1186/1472-6750-7-81.
- Guglielmi L, Denis V, Vezzio-Vié N, Bec N, Dariavach P, et al. (2011) Selection for intrabody solubility in mammalian cells using GFP fusions. *Protein Eng Des Sel* 24: 873–881. doi:10.1093/protein/gzr049.
- Li J, Tibshirani R (2011) Finding consistent patterns: A nonparametric approach for identifying differential expression in RNA-Seq data. *Stat Methods Med Res*. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22127579>. Accessed 28 February 2012.
- Giudicelli V, Duroux P, Ginestoux C, Folch G, Jabado-Michaloud J, et al. (2006) IMGTL/LIGM-DB, the IMGTL(R) comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. *Nucl Acids Res* 34: D781–784.
- Dauvillier S, Merida P, Visintin M, Cattaneo A, Bonnerot C, et al. (2002) Intracellular single-chain variable fragments directed to the Src homology 2 domains of Syk partially inhibit Fc epsilon RI signaling in the RBL-2H3 cell line. *J Immunol* 169: 2274–2283.
- Schneider P (2000) Production of recombinant TRAIL and TRAIL receptor: Fc chimeric proteins. *Methods Enzymol* 322: 325–345.
- Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* 1: 2856–2860. doi:10.1038/nprot.2006.468.
- Fisher AC, DeLisa MP (2009) Efficient Isolation of Soluble Intracellular Single-chain Antibodies using the Twin-arginine Translocation Machinery. *J Mol Biol* 385: 299–311. doi:10.1016/j.jmb.2008.10.051.
- Demo SD, Masuda E, Rossi AB, Thronset BT, Gerard AL, et al. (1999) Quantitative measurement of mast cell degranulation using a novel flow cytometric annexin-V binding assay. *Cytometry* 36: 340–348.
- Asai K, Kitaura J, Kawakami Y, Yamagata N, Tsai M, et al. (2001) Regulation of mast cell survival by IgE. *Immunity* 14: 791–800.
- Kalesnikoff J, Huber M, Lam V, Damen JE, Zhang J, et al. (2001) Monomeric IgE stimulates signaling pathways in mast cells that lead to cytokine production and cell survival. *Immunity* 14: 801–811.
- Ward ES, Güssow D, Griffiths AD, Jones PT, Winter G (1989) Binding activities of a repertoire of single immunoglobulin variable domains secreted from *Escherichia coli*. *Nature* 341: 544–546. doi:10.1038/341544a0.
- Su C, Nguyen VK, Nei M (2002) Adaptive Evolution of Variable Region Genes Encoding an Unusual Type of Immunoglobulin in Camelids. *Mol Biol Evol* 19: 205–215.
- Tanaka T, Lobato MN, Rabbitts TH (2003) Single domain intracellular antibodies: a minimal fragment for direct in vivo selection of antigen-specific intrabodies. *J Mol Biol* 331: 1109–1120.
- Kalesnikoff J, Galli SJ (2008) New developments in mast cell biology. *Nat Immunol* 9: 1215–1223. doi:10.1038/nri.216.
- Genini S, Nguyen TT, Malek M, Talbot R, Gebert S, et al. (2006) Radiation hybrid mapping of 18 positional and physiological candidate genes for arthrogryposis multiplex congenita on porcine chromosome 5. *Anim Genet* 37: 239–244. doi:10.1111/j.1365-2052.2006.01447.x.
- Yea K, Zhang H, Xie J, Jones TM, Yang G, et al. (2013) Converting stem cells to dendritic cells by agonist antibodies from unbiased morphogenic selections. *Proc Natl Acad Sci U S A* 110: 14966–14971. doi:10.1073/pnas.1313671110.
- Xie J, Yea K, Zhang H, Moldt B, He L, et al. (2014) Prevention of Cell Death by Antibodies Selected from Intracellular Combinatorial Libraries. *Chem Biol* 21: 274–283. doi:10.1016/j.chembiol.2013.12.006.
- Ravn U, Gueneau F, Baerlocher L, Osteras M, Desmurs M, et al. (2010) Bypassing in vitro screening—next generation sequencing technologies applied to antibody display and in silico candidate selection. *Nucleic Acids Res* 38: e193. doi:10.1093/nar/gkq789.
- Paz K, Brennan LA, Iacolina M, Doody J, Hadari YR, et al. (2005) Human single-domain neutralizing intrabodies directed against Etk kinase: a novel approach to impair cellular transformation. *Mol Cancer Ther* 4: 1801–1809. doi:10.1158/1535-7163.MCT-05-0174.
- Griffin H, Elston R, Jackson D, Ansell K, Coleman M, et al. (2006) Inhibition of papillomavirus protein function in cervical cancer cells by intrabody targeting. *J Mol Biol* 355: 360–378.
- Cohen PA, Mani JC, Lane DP (1998) Characterization of a new intrabody directed against the N-terminal region of human p53. *Oncogene* 17: 2445–2456.

35. Martineau P, Jones P, Winter G (1998) Expression of an antibody fragment at high levels in the bacterial cytoplasm. *J Mol Biol* 280: 117–127.
36. Tanaka T, Williams RL, Rabbits TH (2007) Tumour prevention by a single antibody domain targeting the interaction of signal transduction proteins with RAS. *EMBO J* 26: 3250–3259. doi:10.1038/sj.emboj.7601744.
37. Miller TW, Zhou C, Gines S, MacDonald ME, Mazarakis ND, et al. (2005) A human single-chain Fv intrabody preferentially targets amino-terminal Huntingtin's fragments in striatal models of Huntington's disease. *Neurobiol Dis* 19: 47–56. doi:10.1016/j.nbd.2004.11.003.
38. Cassimeris L, Guglielmi L, Denis V, Larroque C, Martineau P (2013) Specific In Vivo Labeling of Tyrosinated α -Tubulin and Measurement of Microtubule Dynamics Using a GFP Tagged, Cytoplasmically Expressed Recombinant Antibody. *PLoS ONE* 8: e59812. doi:10.1371/journal.pone.0059812.
39. Caussinus E, Kanca O, Affolter M (2012) Fluorescent fusion protein knockout mediated by anti-GFP nanobody. *Nat Struct Mol Biol* 19: 117–121. doi:10.1038/nsmb.2180.
40. Butler DC, Messer A (2011) Bifunctional Anti-Huntingtin Proteasome-Directed Intrabodies Mediate Efficient Degradation of Mutant Huntingtin Exon 1 Protein Fragments. *PLoS ONE* 6: e29199. doi:10.1371/journal.pone.0029199.
41. Sibler A-P, Nordhammer A, Masson M, Martineau P, Trave G, et al. (2003) Nucleocytoplasmic shuttling of antigen in mammalian cells conferred by a soluble versus insoluble single-chain antibody fragment equipped with import/export signals. *Exp Cell Res* 286: 276–287.
42. Böldicke T (2007) Blocking translocation of cell surface molecules from the ER to the cell surface by intracellular antibodies targeted to the ER. *J Cell Mol Med* 11: 54–70. doi:10.1111/j.1582-4934.2007.00002.x.
43. Rabu C, McIntosh R, Jurasova Z, Durrant L (2012) Glycans as targets for therapeutic antitumor antibodies. *Future Oncol Lond Engl* 8: 943–960. doi:10.2217/fo.12.88.
44. Quintana FJ, Yeste A, Weiner HL, Covacu R (2012) Lipids and lipid-reactive antibodies as biomarkers for multiple sclerosis. *J Neuroimmunol* 248: 53–57. doi:10.1016/j.jneuroim.2012.01.002.
45. Luo J, Emanuele MJ, Li D, Creighton CJ, Schlabach MR, et al. (2009) A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell* 137: 835–848. doi:10.1016/j.cell.2009.05.006.
46. Beronja S, Janki P, Heller E, Lien W-H, Keyes BE, et al. (2013) RNAi screens in mice identify physiological regulators of oncogenic growth. *Nature* 501: 185–190. doi:10.1038/nature12464.
47. Huang S, Hölzel M, Knijnenburg T, Schlicker A, Roepman P, et al. (2012) MED12 controls the response to multiple cancer drugs through regulation of TGF- β receptor signaling. *Cell* 151: 937–950. doi:10.1016/j.cell.2012.10.035.
48. Kontermann RE (2004) Intrabodies as therapeutic agents. *Methods* 34: 163–170.
49. Alvarez RD, Barnes MN, Gomez-Navarro J, Wang M, Strong TV, et al. (2000) A cancer gene therapy approach utilizing an anti-erbB-2 single-chain antibody-encoding adenovirus (AD21): a phase I trial. *Clin Cancer Res* 6: 3081–3087.
50. Swaminathan J, Ehrhardt C (2012) Liposomal delivery of proteins and peptides. *Expert Opin Drug Deliv* 9: 1489–1503. doi:10.1517/17425247.2012.735658.
51. Freund G, Sibler A-P, Desplancq D, Oulad-Abdelghani M, Vigneron M, et al. (2013) Targeting endogenous nuclear antigens by electrotransfer of monoclonal antibodies in living cells. *mAbs* 5: 518–522. doi:10.4161/mabs.25084.
52. Marschall ALJ, Frenzel A, Schirrmann T, Schüngel M, Dubel S (2011) Targeting antibodies to the cytoplasm. *mAbs* 3: 3–16. doi:10.4161/mabs.3.1.14110.
53. D'Agostino R, Belanger A, D'Agostino Jr. R. (n.d.) A suggestion for using powerful and informative tests of normality. *Am Stat* 44: 316–321.