



**HAL**  
open science

# Méthodologie pour l'obtention de structures par cryo-EM de complexes anticorps-antigènes appliquée aux deux protéines sHER2 et NS4

Rémi Ruedas

► **To cite this version:**

Rémi Ruedas. Méthodologie pour l'obtention de structures par cryo-EM de complexes anticorps-antigènes appliquée aux deux protéines sHER2 et NS4. Biologie structurale [q-bio.BM]. Université Paris-Saclay, 2023. Français. NNT : 2023UPASQ059 . tel-04585329

**HAL Id: tel-04585329**

**<https://theses.hal.science/tel-04585329>**

Submitted on 23 May 2024

**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.

# Méthodologie pour l'obtention de structures par cryo-EM de complexes anticorps-antigènes appliquée aux deux protéines sHER2 et NS4

*Methodology for obtaining cryo-EM structures of antibody-antigen complexes applied to the two protein sHER2 and NS4*

**Thèse de doctorat de l'université Paris-Saclay**

École doctorale n°569, Innovation Thérapeutique du Fondamental à l'Appliqué (ITFA)

Spécialité de doctorat : Biologie structurale

Graduate School : Santé et médicament, Référent : Faculté de Pharmacie

Thèse préparée dans l'unité de recherche, **Institute for Integrative Biology of the Cell – I2BC (Université Paris-Saclay, CEA, CNRS)** sous la direction de **Stéphane BRESSANELLI**, directeur de recherche, et la co-direction de **Magali MATHIEU (Sanofi)**, docteur.

Thèse soutenue à Paris-Saclay, le 20 novembre 2023, par

**Rémi RUEDAS**

## Composition du Jury

Membres du jury avec voix délibérative

**Sylvie NESSLER**

Professeure, Université Paris-Saclay

**Valérie BIOU**

Chargée de recherche (HDR), Université Paris Cité

**Pierre-Damien COUREUX**

Professeur, Université Lyon

**Aurélié BERTIN**

Directeur de recherche, CNRS

**Présidente**

**Rapporteur & Examinatrice**

**Rapporteur & Examineur**

**Examinatrice**



**Titre :** Méthodologie pour l'obtention de structures par cryo-EM de complexes anticorps-antigènes appliquée aux deux protéines sHER2 et NS4.

**Mots clés :** Cryo-EM, Trastuzumab, Pertuzumab, Anticorps, Alpharep, Biophysique

**Résumé :** L'analyse en particules uniques (SPA) par microscopie électronique à transmission cryogénique (Cryo-EM) permet aujourd'hui l'analyse structurale à l'état natif d'une large gamme de protéines et de complexes macromoléculaires. Elle est particulièrement intéressante pour les complexes pour lesquels la prédiction de la structure reste difficile, tels que les complexes anticorps-antigènes. Nous couvrons dans ce manuscrit plusieurs points de validation aux étapes pré-microscope, in-microscope et post-microscope. Nous détaillons une large gamme d'analyses biophysiques permettant une description détaillée des propriétés biochimiques des échantillons en solution. À la suite de ces analyses, nous discutons de l'approche optimale de préparation et criblage de grilles de Cryo-EM. La dernière partie est consacrée au traitement des données avec des analyses exhaustives de l'effet du matériau de la grille sur l'échantillon. Nous exposons une application pratique des logiciels actuels pour faire face à l'hétérogénéité conformationnelle continue. Ces travaux ont permis l'obtention de la première carte expérimentale au-delà de 3 Angströms de résolution du complexe ternaire flexible "HTP" de 162 kDa entre le récepteur du facteur de croissance épidermique humain 2 (HER2) et les fragments de liaison à l'antigène (Fab) de deux anticorps thérapeutiques distincts, le pertuzumab et le trastuzumab.

Notre carte mieux définie que les précédentes, à permis de mettre en lumière une interaction précédemment négligée qui pourrait expliquer l'effet synergique des anticorps.

Le troisième chapitre reprend l'approche développée avec le complexe HTP en l'appliquant à la protéine non structurale de norovirus NS4. Cette partie apporte les premiers résultats exhaustifs des propriétés biochimique de la protéine NS4. Elle expose également les limitations de l'utilisation des anticorps monoclonaux comme outils d'analyse structurale. Enfin il est montré comment nous avons réussi à obtenir un catalogue diversifié de protéines artificielles "alpharep" ciblant NS4 ainsi que les premières applications possibles.

**Title :** Methodology for obtaining cryo-EM structures of antibody-antigen complexes applied to the two protein sHER2 and NS4

**Keywords :** Cryo-EM, Trastuzumab, Pertuzumab, Antibodies, Alphareps, Biophysic

**Abstract :** Single particle analysis (SPA) from cryogenic transmission electron microscopy (Cryo-EM) nowadays allows structural analysis in the native state of a broad range of proteins and macromolecular complexes. It is particularly attractive for complexes for which structure prediction remains intractable, such as antibody-antigen complexes. In the second chapter of this manuscript, we cover several validation points at the pre-microscope, in-microscope and post-microscope stages for the studies of such complexes by cryo-EM. We detail a wide range of biophysical analyses allowing a detailed description of the biochemical properties of samples in solution. Following these analyses, we discuss the optimal approach for preparing and screening Cryo-EM grids. The last part is devoted to data processing, with exhaustive analyses of the effect of the grid material on the sample. We also present a practical application of current software to deal with continuous conformational heterogeneity. This part shows how we manage to obtain the first experimental map below 3 Angstroms resolution of the flexible ternary complex "HTP" between the human epidermal growth factor receptor 2 (HER2) and the antigen-binding fragments (Fab) of two distinct therapeutic antibodies, pertuzumab and trastuzumab.

Our map, which is better defined than previous ones, has brought to light a previously neglected interaction that could explain the synergistic effect of the antibodies.

The third chapter takes up the approach developed with the HTP complex by applying it to the non-structural norovirus protein NS4. This section provides the first comprehensive results on the biochemical properties of the NS4 protein. It also outlines the limitations of using monoclonal antibodies as structural analysis tools. Finally, it shows how we have succeeded in obtaining a diversified catalogue of artificial proteins "alpharep" targeting NS4, as well as the first possible applications.

## Remerciements

Tout d'abord, je tiens à exprimer ma gratitude pour mon directeur de thèse Stéphane Bressanelli. Ce projet n'aurait pas été ce qu'il est sans tes conseils et ta supervision. Tu m'as donné cette chance d'entrer dans le monde de la cryo-EM, et je te suis infiniment reconnaissant pour cela. Il y a de nombreux points sur lesquels j'aimerais sincèrement te remercier. Bien sûr, il y a l'aspect scientifique. Tu as toujours pris le temps de transmettre tes connaissances avec patience, pédagogie et modestie. J'aimerais aussi souligner ta volonté de porter tes thésards vers le haut et à quel point tu nous fais confiance. Je me souviendrai de ta passion pour la recherche, qui n'a cessé de nourrir la mienne. Enfin, je me souviendrai également de ta démonstration par l'expérience sur ta personne, tout au long de ces trois années, que le sucre n'est pas aussi mauvais pour la santé qu'on le dit.

Je veux aussi remercier ma codirectrice Magali Mathieu de m'avoir donné cette opportunité d'entrer dans le monde de l'industrie chez Sanofi. C'est grâce à toi que j'ai eu cette chance de découvrir la R&D du monde pharmaceutique. Merci pour tes conseils tout au long du projet et pour ton aide dans la relecture de ce manuscrit. Merci de m'avoir fait découvrir Paris le temps d'une longue randonnée et pour cette belle impression 3D.

Je ne peux pas omettre de remercier Sonia Fieulaine qui, tel Atlas en mythologie, porte le labo sur ses épaules. Merci de m'avoir si souvent laissé venir te demander conseil à propos de mon western-blot pas génial, mon profil d'élution chaotique, et d'avoir toujours résolu l'énigme "ils sont rangés où ?". Merci pour tes précieux conseils scientifique et humain qui m'ont beaucoup aidé. Merci de m'avoir transmis tes valeurs et ta rigueur scientifique. Merci de m'avoir partagé tous les secrets du labo, surtout celui de l'oiseau.

Le plus électronique des mercis à toutes ces personnes qui ont largement contribué à ma formation en cryo-EM. Tout d'abord, il y a Jean-Marie Winter. Merci de m'avoir transmis une infime partie ton savoir et de m'avoir parfois éclairé, parfois noyé au cours de nos longues conversations sur la microscopie électronique. Merci pour ton dévouement qui a été crucial à ce projet. Merci aussi à Ana Andreea-Artenis, pour sa confiance, ses conseils et son soutien. C'est avec toi que j'ai commencé la cryo-EM, et tu m'as soutenu à chaque fois que tu en avais l'occasion, merci. Merci Laura Pieri pour ces longs échanges et rires sur les concepts théoriques de cryo-EM, parfois assez abstraits. Enfin merci à toi Malika Ould Ali, qui m'a accompagné et initié sur les fragiles microscopes tecnai. Toi et Ana aurez bien voulu que j'en casse un pour le remplacer par un Glacios, mais vous m'avez trop bien formé, dommage.

Le plus fou, amical, et sincère merci à tous les membres de l'équipe SB2SM/IMAPP qui m'ont accueilli à bras ouverts, Maité Paternostre (le rayon de soleil), Marion Babot (la femme invisible), Virginie Gervais (la cycliste), Marie-Hélène Le Du (la joie de vivre) et Yves Boulard (le Breton). C'est avec vous que j'ai passé la majeure partie de mon temps ces trois dernières années avec beaucoup de plaisir. En montant à Paris, j'ai perdu le soleil de Montpellier et les

montagnes de Grenoble, mais j'ai trouvé beaucoup d'humour, de bienveillance et de respect. Je n'aurais pas pu rêver mieux et si c'était à refaire, je recommencerais sans hésiter.

Merci à Thibault Tubiana, je sais que tu aurais voulu arriver plus haut dans les remerciements, mais priorité aux manips, et tu es bio-informaticien. Merci pour ta compagnie dans le bureau, tes conseils, ton aide et les moments de rires que tu nous offres. Tu m'as fait découvrir la plupart des outils bio-informatiques et d'illustration qui sont dans cette thèse et qui me suivront tout au long de ma carrière. Et sache que ce paragraphe n'a pas été rédigé par chat-GPT.

Le plus grand merci à tous les membres de l'équipe de modélisation et d'ingénierie des protéines : Philippe Minard, Agathe Urvoas-Cisse, Marie-Béatrice Valerio-Lepinec, Stéphanie Thébault, Agnès Mesneau et Antoine Hivert, Melissa Jin, et Elena Brusoni de m'avoir accueilli avec autant d'entrain et de patience. J'ai pris un vrai plaisir à travailler avec vous, autant du point de vue de la richesse scientifique que pour la jovialité et l'humanité de votre équipe. Il s'en est fallu de peu que je ne reparte plus. Au plaisir d'interagir avec vous à nouveau à l'avenir.

Merci aux membres du Jury Sylvie Nessler, Sophie Bioux, Pierre-Damien Coureux et Aurélie Bertin d'avoir accepté d'expertiser mes travaux de thèse et d'avoir si brillamment et habilement animé le débat lors de ma soutenance.

Merci à tout le département BSB de Sanofi pour votre accueil et pour votre aide. Un merci tout particulier à Valérie Steiner, Anand Kumar, Yulia Dubianok, Gilles Haussy, Armelle Buzy et Thomas Bertrand pour vos précieux conseils. Et merci à Alexey Rak de m'avoir accueilli dans le service. J'ai apprécié être en votre compagnie à Vitry.

Merci à mon comité de suivi de thèse Aurélie Bertin et Ludovic Sauguet d'avoir pris le temps de m'écouter et de me conseiller.

Danke schön à Volker Lohmann, Marit Lingemann, et Robin Veenstra pour tous ces riches échanges scientifiques. C'est d'abord avec beaucoup d'appréhension que j'attendais nos réunions biennuelles. Mais l'émotion s'est vite transformée en enthousiasme lorsque j'ai réalisé avec quelle sympathie, respect et rigueur se déroulaient nos échanges. Vous m'avez parfaitement illustré l'importance de la collaboration, qui amène de nouvelles idées et un boost de motivation.

Mon plus sérieux merci à Ines Li De La Sierra-Gallay, Paloma Fernandez et Zahra Marouf de m'avoir invité à travailler avec vous sur vos jeux de données. Vous avez joué un rôle crucial dans mon expérience en cryo-EM.

Mon plus fidèle merci à Kalouna Kra et Adrien Royet. Nous aurons vécu l'épreuve de la thèse ensemble avec ses hauts et ses galères. Cela fait de nous des frères de thèse ! Nos pauses

goutées vont me manquer.

Mon plus studieux merci à Karim Ben-Ali-Gacem avec qui j'ai partagé le Glacios. On a élevé des grilles ensemble et rare sont ceux qui peuvent en dire autant.

Mon plus chaleureux merci à Camille Samson, François Vallée et Oriane Frances qui sont des personnes exceptionnelles que je suis heureux d'avoir rencontrées. Et n'oubliez pas, nous sommes liés par le karma 2.

Mon plus drôle merci à Murielle Seif El Dahan pour la visite du musée de l'illusion.

Un merci tout particulier à Françoise Jean-Jean, qui m'a soutenu pendant les années difficiles du collège et du lycée. C'est grâce à votre aide que je peux aujourd'hui m'épanouir dans cette science que j'aime tant.

Le plus aimant remerciement à ma famille. Ma mère et mon père qui m'ont toujours soutenu et qui ont fait tant d'efforts pour me permettre d'en arriver là. Merci à ma sœur Célia Ruedas dont je m'inspire, qui me motive et dont je suis si fier. Merci à mes oncles, tantes, cousin et grands-parents qui m'ont soutenu. Même loin des yeux, vous restez très proche du cœur.

Enfin, un dernier merci à mes plus vieux amis. Merci à toi Dr. Pierre Mougin avec qui je partage mes plus vieux et mes beaux souvenirs, depuis les JSP jusqu'aux montagnes, vivement la prochaine aventure ! Merci à toi Marine Roche qui est à mes coté depuis 14 années, une amie d'exception toujours disponible pour donner de sages conseils. Merci à Claire Saragaglia pour tes conseils, ton écoute, ta bienveillance et pour si bien organiser nos voyages. Merci à ma partenaire de concert de métal depuis 6 ans Charline Gauthier. Merci à Pauline et Corentin Estoudre-Aubert que nos choix de vie ont éloignés mais qui garde une place de choix dans ma vie. Vous tous m'avez aidé à garder la tête sur les épaules et le moral, merci.





## Table des matières

Abréviations .....	12
<b>1 Introduction</b> .....	16
1.1 La microscopie électronique à transmission (TEM) .....	18
1.1.1 De l'électron à la microscopie électronique .....	18
1.1.1 L'architecture d'un microscope .....	19
1.1.2 Traitement des images de l'analyse en particule unique (SPA) par cryo-EM .....	31
1.2 HER2-Trastuzumab-Pertuzumab.....	37
1.2.1 La famille des récepteurs transmembranaires ErbB.....	37
1.2.2 L'origine et le rôle des anticorps.....	39
1.2.3 L'origine de la diversité des anticorps monoclonaux.....	41
1.2.4 Les méthodes de productions et l'humanisation des anticorps.....	42
1.2.5 Les dérivés d'anticorps .....	43
1.2.6 Les applications des anticorps.....	44
1.2.7 Trastuzumab.....	46
1.2.8 Pertuzumab.....	47
1.2.9 Bilan des structures disponibles.....	47
1.3 Principes biophysiques .....	48
1.3.1 Nano-DSF .....	48
1.3.2 DLS.....	50
1.3.3 SEC-SLS .....	51
1.3.4 Micro-Scale Thermophoresis .....	52
1.3.5 Photométrie de masse.....	53
1.3.6 Spectrométrie de masse MALDI-TOF.....	54
1.4 Alphareps.....	55
1.4.1 Alphareps origine .....	55
1.4.2 L'architecture des alphareps.....	56
1.4.3 Design de la bibliothèque et création de la diversité .....	58
1.4.4 Phage display .....	59
1.5 NS4.....	60
<b>2 Structure du complexe sHER2-Pertuzumab-Trastuzumab Fabs</b> .....	62
2.1 Pre-microscope : Une vue biochimique et biophysique de l'échantillon.....	64

2.1.1	Matériel et méthodes : purification, analyses biophysiques.....	64
2.1.2	Résultats : une vue exhaustive du comportement du complexe par les méthodes biophysique.....	69
2.1.3	Conclusion et discussion: un échantillon stable et homogène en solution .....	80
2.2	In-microscope : Préparation de grilles et approche optimisée pour un screening de grilles efficient. La quête de la grille ultime.....	82
2.2.1	Matériel et méthode : la préparation des grilles et le screening sur le microscope.....	82
2.2.2	Résultats : un effet grille dépendant .....	85
2.2.3	Conclusion et discussion : la stratégie de screening de grilles pour une collecte dans des conditions optimales .....	90
2.3	Post-microscope : Le traitement des données.....	94
2.3.1	Matériel et méthode : de la collecte au traitement des données .....	94
2.3.2	Résultats.....	105
2.3.3	Conclusion et discussion : nouvelle structure du complexe HTP et analyse de la flexibilité continue .....	131
<b>3</b>	<b>La protéine de non structurale 4 (NS4) de norovirus.....</b>	<b>134</b>
3.1	Caractérisation biochimique, biophysique et prédictions bio-informatiques de la protéine NS4 .....	136
3.1.1	Matériel et méthode : clonage, expression et purification de NS4 .....	136
3.1.2	Résultats : Une vue exhaustive de NS4.....	142
3.1.3	Conclusion et discussion : une meilleure compréhension de NS4 par la biophysique et les outils bioinformatiques.....	150
3.2	Des Fabs pour une structure par cryo-EM de la protéine NS4 .....	152
3.2.1	Matériel et méthode : purification des mABs et des Fabs.....	152
3.2.2	Résultat : des conditions biochimiques incompatibles avec l'utilisation de Fab.....	159
3.2.3	Conclusion et discussion : les dérivés d'anticorps ne sont pas les outils idéaux pour l'étude structurale .....	168
3.3	Les Alphareps dirigées contre la protéine NS4.....	170
3.3.1	Matériel et méthode.....	170
3.3.2	Résultats.....	179
3.3.3	Conclusion et perspectives : un catalogue d'alpharep diversifié disponible .....	186
<b>4</b>	<b>Discussion Générale.....</b>	<b>190</b>

<b>5</b>	<b>Références</b> .....	194
<b>6</b>	<b>Annexes</b> .....	206
6.1	Alignements des différent génogroupes de NS4.....	206
6.1.1	Alignement des séquences de GI .....	206
6.1.2	Alignement des séquences de GII.4.....	207
6.1.3	Alignement des séquences de GII .....	207
6.1.4	Alignement des différentes séquences des différents génogroupes .....	208
6.2	Article soumis à IUCrJ le 23/08/2023.....	209



## Abréviations

$\mu\text{M}$  = Unité de micromoles/litre

ADN = Acide désoxyribonucléique

CDR = Régions déterminant la complémentarité (complementary determining region)

Cryo-EM = Microscopie électronique cryogénique

DDM = n-Dodecyl-beta-D-maltoside

DLS = Diffusion dynamique de la lumière (Dynamic light scattering)

DMSO = Diméthylsulfoxyde

DSF = Fluorimétrie différentielle à balayage (Differential scanning fluorimetry)

DTT = Dithiothreitol

*e.coli* = Escherichia coli

EDTA = Acide éthylènediaminetétraacétique

Eer = electron-event representation

ELISA = Méthode immuno-enzymatique (Enzyme-Linked Immuno Sorbent Assay)

eV = Électron volt

Fab = Fragment liant l'antigène (Fragment antigen binding)

FPLC = Chromatographie liquide rapide pour protéine (Fast protein liquid chromatography)

Fps = Images par seconde (Frames per second)

HRP = Peroxydase de raifort (horseradish peroxidase)

IDDT = Test de différence locale (Local distance difference test)

IgG = Immunoglobuline G

IPTG = Isopropyl  $\beta$ -D-1-thiogalactopyranoside

Kd = Constante de dissociation

kDa= Kilodalton

LB = Milieu de lysogénie

LRR = Répétition riche en leucine

mAB = Anticorps monoclonal (monoclonal antibody)

MALDI-TOF = désorption-ionisation laser assistée par matrice (Matrix assisted laser desorption ionisation)

MES = Acide 2-(N-morpholino)éthanosulfonique

MST = Analyse par thermodiffusion micro-échelle (microscale thermophoresis)

MW = Poids moléculaire en grammes/moles

Ni-NTA = Acide nitrilotriacétique de nickel

NS4 = Protéine non-structurale de norovirus 4

OD = Densité optique

ORF = Cadre de lecture ouvert (Open reading frame)

PAGE = Gel électrophorèse en polyacrylamide

PCR = Réaction de polymérase en chaîne

PDI = Indice de polydispersité

pH = Potentiel hydrogène

RCA = Amplification de type "rolling circle" ou Réplication circulaire de l'ADN

Rh = Rayon hydrodynamique

SDS = Sodium dodecyl sulfate

SEC = Chromatographie d'exclusion en taille (size exclusion chromatography)

SEC-SLS = Chromatographie d'exclusion en taille et diffusion statique que la lumière (Size exclusion chromatography and static light scattering).

SPA = Analyse par particule unique (Single particule analysis)

STEM = Microscopie électronique à transmission à balayage (Microscopie

électronique à transmission à balayage)

TEM = Microscopie électronique à transmission

T<sub>m</sub> = Température de fusion

TOF = Temps de vols (Time of flight)

TPR = Tetratricopeptide repeat

Tris = 2-amino-2-hydroxyméthylpropane-1,3-diol

UV = Ultraviolet

β-OH = Betamercaptoéthanol

PVDF = Polyvinylidene difluoride





# 1 Introduction

Dans cette introduction, je vais d'abord décrire la méthode principale qui m'a permis d'étudier la structure des protéines durant ma thèse, la cryo-microscopie électronique. Je donnerai ensuite les éléments de contexte sur le complexe protéique HTP qui a été mon objet d'étude principal. Je décrirai ensuite les méthodes biophysiques utilisées dans le manuscrit. Enfin, je présenterai succinctement les alphareps et la protéine non structurale de norovirus NS4.

La microscopie électronique à transmission

HER2-Trastuzumab-Pertuzumab

Principes biophysiques

Les Alphareps

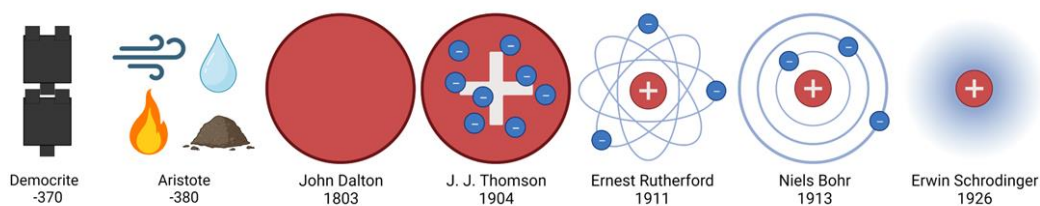
NS4



## 1.1 La microscopie électronique à transmission (TEM)

### 1.1.1 De l'électron à la microscopie électronique

Le concept d'atome prend vie au Ve siècle avant J.-C. par Leucippe et Démocrite, marquant le début de l'ère de l'atomisme philosophique. L'atome prend le sens de particules invisibles et indivisibles qui s'assemblent entre elles pour former la matière. Cette théorie reste en arrière-plan durant deux millénaires, éclipsée par la théorie des quatre éléments d'Aristote. Mais l'atome alimente le courant matérialiste qui se fraie un chemin jusqu'à la révolution chimique du XVIIIe siècle et aux travaux du chimiste Antoine Lavoisier, qui amènent à la loi de conservation qui porte son nom. Au début du XIXe siècle, John Dalton remet au goût du jour la théorie de l'atome et propose un modèle introduisant plusieurs notions encore vraies aujourd'hui. Un siècle après, ce sont les travaux de Joseph John Thomson qui introduisent le concept d'électrons qui composent l'atome, via l'expérience des tubes cathodiques. Une expérience très similaire au fonctionnement des canons à électrons que l'on retrouve dans un microscope électronique moderne. Cette théorie marque le début de l'explosion fulgurante des connaissances sur l'atome, puisqu'en 1911, le modèle de Rutherford est adopté et dissocie les électrons du noyau. Ce modèle est rapidement complété par Niels Bohr, qui propose de classer les électrons par couche d'énergie. Enfin, Louis de Broglie postule la propriété onde-particule des électrons en 1924, ainsi qu'une équation pour calculer la longueur d'onde d'un faisceau de particules, qui est inversement proportionnelle à la vitesse des particules qui le composent. Il se base sur la démonstration faite par Einstein quelques années plus tôt, qui introduit l'idée de la dualité onde-corpuscule de la lumière. Einstein reçoit le prix Nobel de physique en 1922 pour ces travaux. Enfin, le modèle atomique est encore perfectionné par Erwin Schrödinger en 1926, qui introduit des concepts quantiques et définit les électrons comme un nuage de probabilité de présence.



**Figure 1.1.1-1.** Les modèles de l'atome au fil de l'histoire. Notons que à par le modèle d'Aristote l'évolution de l'atome est une amélioration du concept par l'expérience et non pas une refonte du concept.

Le premier microscope électronique à transmission expérimental est présenté par le docteur en électrotechnique Max Knoll et son étudiant Ernst Ruska en 1933. Ils se basent sur les expériences des tubes cathodiques, notamment pour la conception d'un canon. La longueur d'onde de de Broglie motive la création d'un tel instrument, car un faisceau composé d'électrons a une longueur d'onde de l'ordre du picomètre, ce qui en fait le rayonnement théoriquement le plus résolutif. De plus, l'électron étant chargé, le faisceau peut être

concentré et manipulé comme il est possible de le faire pour un microscope optique, mais à l'aide de lentilles électromagnétiques. Cependant, un tel rayon est particulièrement destructeur pour l'échantillon, surtout les échantillons biologiques principalement composés d'atomes légers. Enfin, une autre grosse limitation est le faible pouvoir pénétrant du faisceau. C'est sur ces limitations que Jacques Dubochet, Joachim Franck et Richard Henderson ont travaillé pour grandement développer la microscopie électronique cryogénique (Cryo-EM) sur les échantillons biologiques. Ils ont reçu le prix Nobel de Chimie en 2017 pour ces travaux.

## 1.1.2 L'architecture d'un microscope

### 1.1.1.1 Le canon à électron

Le canon à électrons est le premier élément du microscope. Il consiste en l'application d'un courant élevé dans un filament, ce qui conduit à l'émission d'électrons de sa surface en fonction de l'énergie de la fonction de travail du matériau. Le canon est couvert par une capsule appelée le cylindre de Wehnelt, qui permet une première focalisation du faisceau. Un champ électrique appliqué par une pile d'accélération permet d'accélérer le faisceau d'électrons à des vitesses proches de celle de la lumière (~77 %). Depuis le canon et jusqu'au détecteur, l'intégralité de la colonne se trouve dans un vide très poussé.

Les équations de de Broglie et la limite de diffraction d'Abbe permettent d'estimer la résolution théorique d'un faisceau d'électrons. À partir de l'équation de de Broglie, nous pouvons extraire la longueur d'onde du rayonnement d'électrons :

$$\lambda = \frac{h}{mv}$$

Avec  $\lambda$  = longueur d'onde,  $h$  la constante de plank ( $6,6 \cdot 10^{-34} \text{ kg.m}^2.\text{s}^{-1}$ ),  $m$  = la masse de la particule et  $v$  = la vitesse de l'onde.

La vitesse d'un électron étant de  $2,3 \cdot 10^8 \text{ m.s}^{-1}$  à un champ de 300kV, et la masse de l'électron de  $9,1 \cdot 10^{-31} \text{ kg}$ , la longueur d'onde est d'environ 3 pm. En réalité, à ces vitesses très proche de celle de la lumière, il faut corriger la masse de l'électron pour prendre en compte l'effet relativiste. La masse de l'électron sous un champ de 300kV est de  $1.45 \cdot 10^{-30} \text{ kg}$  et donc la longueur d'onde de 2 pm (Williams and Carter 2009; Russell et al. 2019; JEOL 2023)

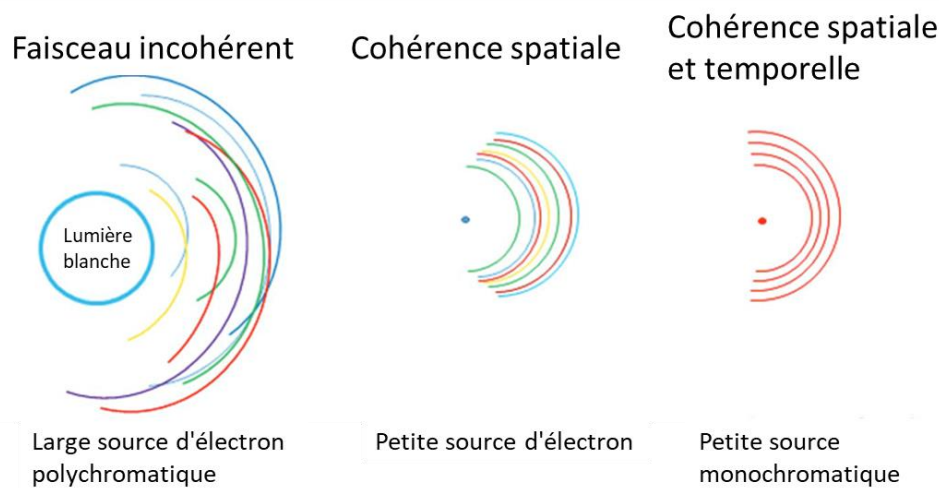
Le critère d'Abbe permet d'estimer la résolution théorique d'un microscope en fonction de la longueur d'onde du rayonnement (Penczek 2010) :

$$d = 0,61 \frac{\lambda}{n \sin \theta}$$

Avec  $\lambda$  = la longueur d'onde du faisceau et  $n \sin \theta$  = l'expression de l'ouverture optique, elle varie avec le diaphragme C2 (C2 aperture) mais peut être approximée à ~0,01 pour un microscope électronique (Penczek 2010).

La résolution maximale théorique d'un microscope électronique est donc de 1,2 pm, c'est-à-dire subatomique. En réalité la résolution est limitée par les composants du microscope notamment les lentilles et la caméra.

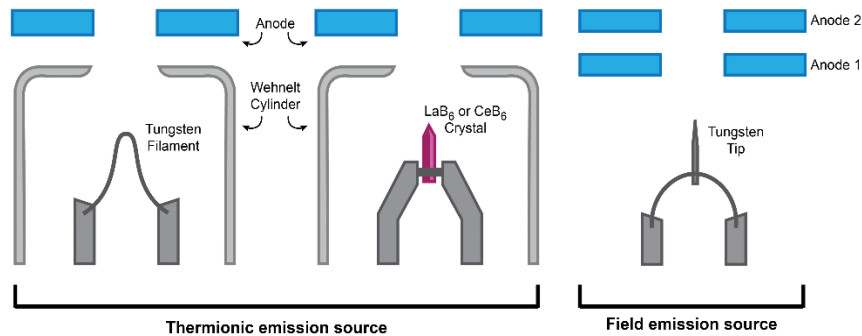
Il existe différents types de canons à électron. Pour bien comprendre ce qui les distingue, il faut introduire le concept de cohérence du faisceau d'électrons. La cohérence spatiale correspond à la divergence des directions des électrons, tandis que la cohérence temporelle correspond à la divergence de la vitesse des électrons. Un faisceau à faible cohérence spatiale entraîne un dédoublement (ou plus) de l'image formée, tandis qu'une faible cohérence temporelle impactera la précision de la focalisation (figure 1.1.2.1-1). Dans les deux cas, la qualité de l'image est directement altérée (Franken et al. 2020). Le cylindre de Wehnelt apporte une première amélioration de la cohérence spatiale.



**Figure 1.1.2-1.** Illustration adaptée de Franken et al. 2020 (Franken et al. 2020) À gauche, un faisceau polychromatique produit par une large source. Au centre, un faisceau polychromatique mais issu d'une source infiniment petite, lui conférant une cohérence spatiale parfaite. À droite, un faisceau monochromatique produit par une petite source ayant donc à la fois une cohérence spatiale et temporelle. La microscopie électronique à transmission a besoin d'un faisceau le plus proche possible de l'illustration de droite (cohérent spatialement et temporellement).

Historiquement les premiers canons sont les filaments thermo-ioniques qui génèrent des électrons en étant chauffés à très haute température. Les filaments de tungstène sont les plus anciens, ils sont particulièrement rustiques, peu chers, et produisent une grande quantité d'électrons. En revanche ils ont une très faible cohérence (spatiale et temporelle) et une faible durée de vie. On trouve ensuite les canons dont la pointe est un cristal d'hexaborure de lanthane ( $\text{LaB}_6$ ) qui ont une meilleure cohérence et une durée de vie plus longue. Enfin, on trouve les filaments dits à émission de champ (FEG) qui sont formés par une très fine pointe de tungstène et se basent sur l'effet tunnel. Ils sont aujourd'hui les canons qui produisent la plus grande quantité d'électrons (brillance) avec la plus haute cohérence et la plus longue

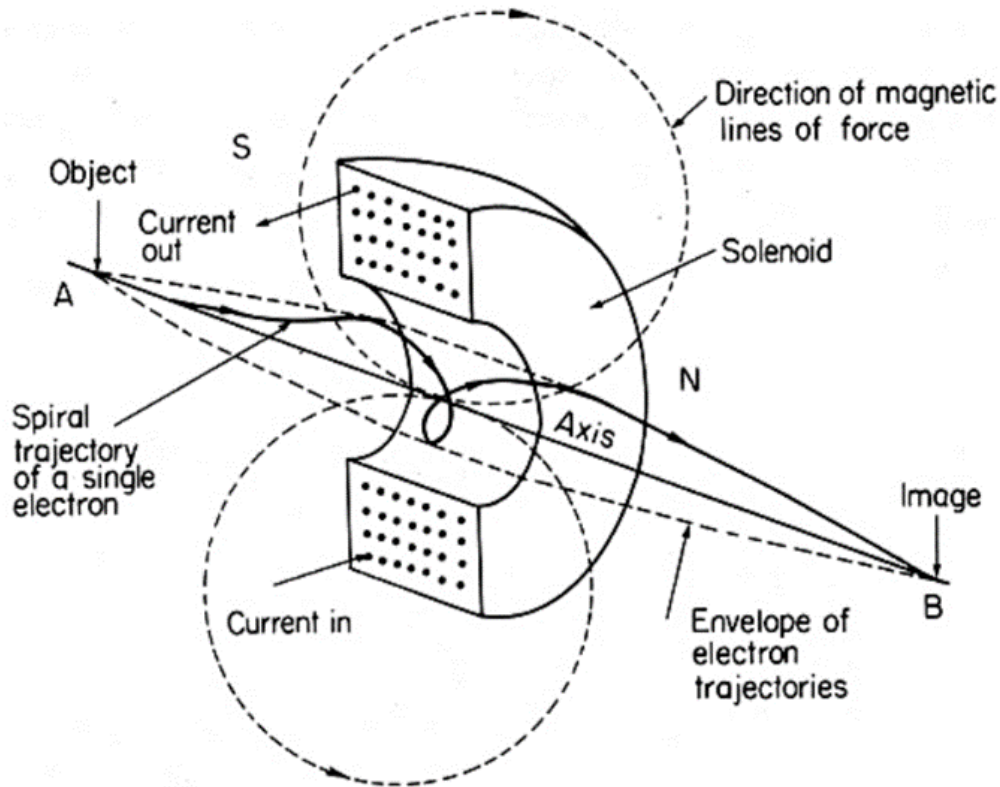
durée de vie (figure 1.1.2.1-2).



**Figure 1.1.2.1-2.** Images reprise de ThermoFisher Scientific, par Alex Ilitchev (Ilitchev et al. 2019). Les différents types de canons à électron. À gauche, les canons thermo-ioniques qui fonctionnent par élévation de la température. Cette famille comprend les canons composés d'un filament de tungstène et ceux avec un cristal d'hexaborure de lanthane (LaB<sub>6</sub>) ou de cérium (CeB<sub>6</sub>). À droite, les filaments à émission de champ composés d'une fine pointe en tungstène. Ceux-ci se divisent en deux sous-groupes, avec ou sans élévation de la température (Cold-FEG).

### 1.1.1.2 Les lentilles en microscopie électronique

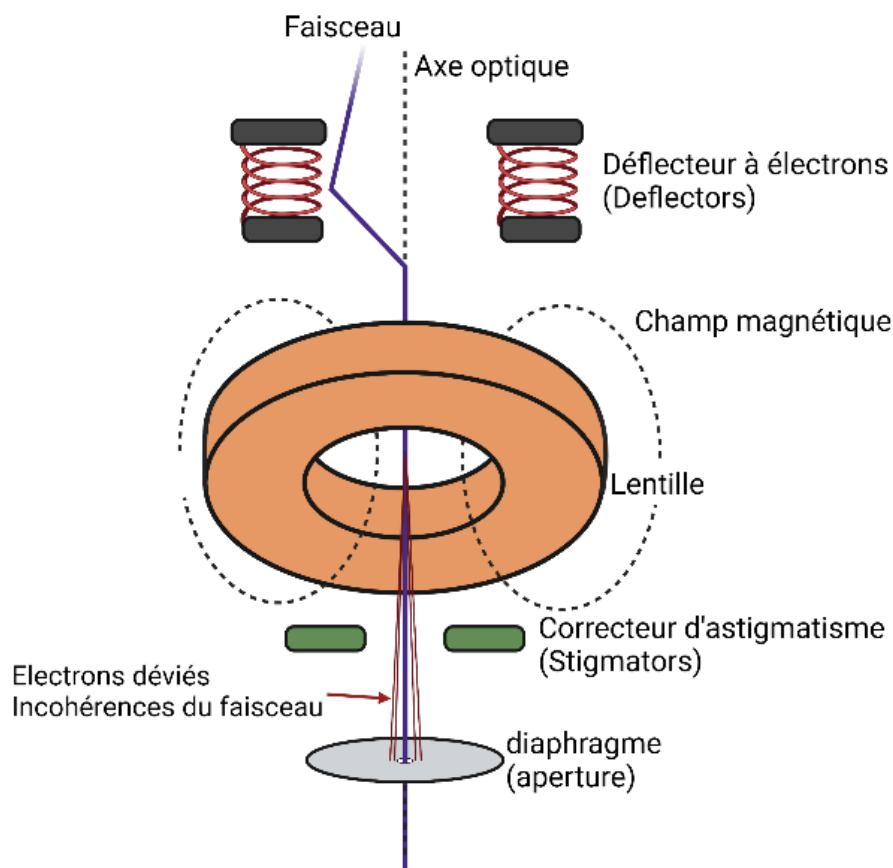
Contrairement aux rayons X et à la lumière, les électrons sont des particules chargées. Leur trajectoire ne peut pas être déviée par une lentille physique en verre en raison de leur faible pouvoir de pénétration et de leur forte interaction avec la matière. En revanche, ils peuvent être contrôlés par des lentilles et d'autres systèmes électromagnétiques. La lentille d'un microscope électronique est un champ magnétique créé par un courant appliqué dans un filament toroïdal que les électrons vont traverser avec une trajectoire hélicoïdale. Cette trajectoire en spirale est responsable de la rotation de l'image lors des changements de grossissement (Figure 1.1.2.2-1).



**Figure 1.1.2.2-1.** Illustration tirée du cours du Pr. Grant Jensen "Getting started in Cryo-EM". Schéma de l'anatomie d'une lentille électromagnétique et de la trajectoire d'un électron unique. L'électron démarre du point A et traverse la bobine solénoïde qui crée un champ de force proportionnel au courant qui la traverse. Le faisceau d'électrons incident subit des forces qui amènent les trajectoires de chaque électron à converger vers le point focal B.

Les lentilles d'un microscope électronique sont des systèmes très imparfaits qui introduisent de nombreuses aberrations. Les principales sont l'aberration sphérique, chromatique, et l'astigmatisme. L'aberration sphérique est l'incapacité à uniformément focaliser les faisceaux incidents, selon qu'ils soient plus ou moins proches de l'axe optique. L'aberration chromatique est provoquée par une focalisation différente selon la longueur d'onde de l'électron incident. Enfin, l'astigmatisme est lié à une disparité de la focalisation le long d'un axe de la lentille. Les aberrations sphériques et chromatiques sont corrigées par le diaphragme (aperture) placé en aval de chaque lentille. L'astigmatisme est corrigé par des quadripôles placés après la lentille (Stigmators) (figure 1.1.2.2-2).





**Figure 1.1.2.2-2.** Système de lentille d'un microscope électronique. Chaque lentille est associée à des déflecteurs à électrons placés en amont de la lentille. Ces bobines dévient le faisceau pour qu'il entre dans la lentille au centre et parallèle à l'axe optique. En aval de la lentille, on trouve des correcteurs d'astigmatismes (Stigmators). Ces quadripôles corrigent l'astigmatisme introduit par la lentille. Enfin, le bloc est terminé par un diaphragme (aperture) qui bloque les électrons déviés de l'axe optique et qui n'ont pas été correctement convergés à cause d'une mauvaise cohérence spatiale (trajectoire) ou temporelle (énergie).

### 1.1.1.3 Chambre de l'échantillon

C'est l'emplacement où l'échantillon sera introduit pour être observé. C'est ici que se passe tout ce qui est en lien avec le déplacement le long de l'objet, en l'occurrence une grille pour la cryo-EM. Il est possible de bouger physiquement sur le plan de la grille (x/y), mais également de haut en bas. Il est aussi possible, à l'aide de déflecteurs, de bouger le faisceau. C'est ici qu'intervient la notion de hauteur eucentrique, qui est assimilable à la position de l'échantillon en Z pour laquelle la zone imagée est placée sur le point de pivot (pivot point). Ainsi, l'image projetée est toujours la même lorsque le support est basculé/incliné autour de l'axe X du plan de la grille.

Parlons de l'échantillon en cryo-TEM, et plus particulièrement en analyse de particule unique

(SPA). Encore une fois, les propriétés des électrons nous contraignent à travailler avec un échantillon très fin pour espérer obtenir une projection de l'image de haute qualité. L'épaisseur de l'échantillon est directement liée à la notion de libre parcours moyen (mean free path), qui correspond à la distance que parcourt une particule dans un milieu avant d'entrer en collision avec une autre particule de ce milieu. Cette valeur oscille autour des 10 nm en TEM (Williams and Carter 2009) et dépend de la composition et de la densité atomique de l'échantillon ainsi que de l'énergie des électrons (puissance du microscope de 200 kV ou 300 kV). Il est mentionné dans la littérature de travailler avec un échantillon épais de 50 à 150 nm pour limiter au maximum les différentes diffusions et donc atteindre la meilleure résolution possible (Passmore and Russo 2016; Rice et al. 2018; Neselu et al. 2023).

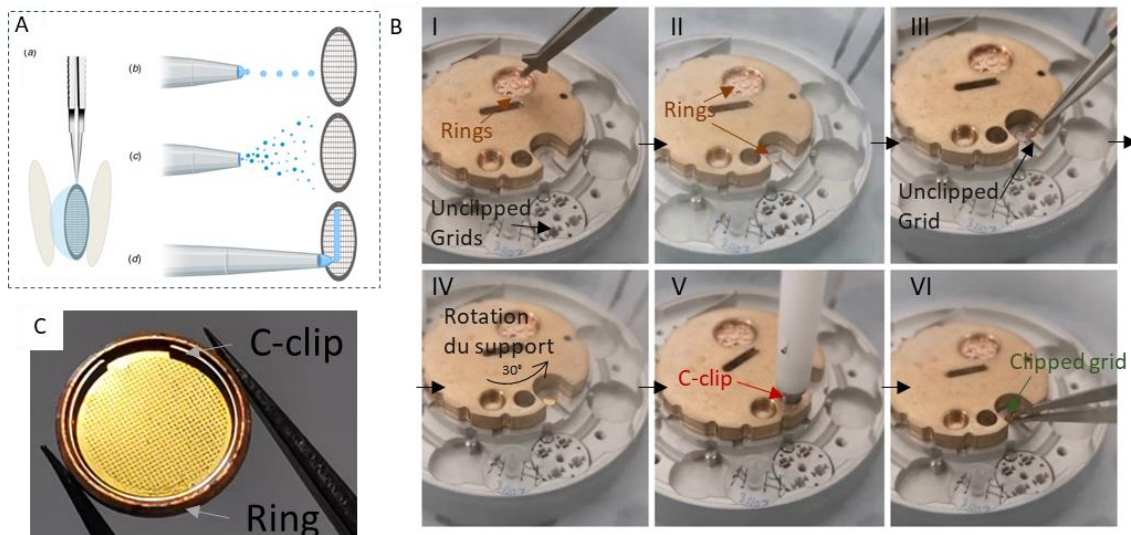
De plus, le faisceau d'électron est très agressif avec la matière et irradie l'échantillon très rapidement (Baker and Rubinstein 2010). C'est la raison pour laquelle il est nécessaire de travailler sur un échantillon vitrifié dans une glace amorphe. Pour réussir à se placer dans de telles conditions pour l'étude de protéines en solution, l'échantillon est déposé sur une grille préalablement traitée dans une chambre à plasma afin de rendre la surface hydrophile. Le surplus de liquide est absorbé avec du papier Whatman, puis la grille est rapidement vitrifiée, le plus souvent par vitrification plongeante dans de l'éthane liquide à l'aide d'un vitrobot, par exemple. L'éthane est un meilleur conducteur thermique et réduit fortement l'effet Leidenfrost du fait de sa température d'ébullition (-88,5 °C), qui est plus élevée que celle de l'azote liquide (-196 °C). Ainsi, l'eau est congelée sous une forme amorphe. D'autres méthodes de préparation d'échantillons existent, le vitrobot étant une méthode de préparation particulièrement peu reproductible. (D'Imprima and Kühlbrandt 2021). Cette étape de vitrification introduit une nouvelle limitation de la cryo-EM, qui concerne la manière dont les particules en solution s'organisent (Kampjut et al. 2021). Il est très courant d'observer que les protéines s'orientent de manière préférentielle, exposant systématiquement une face à l'interface air-eau. Il est également fréquent que les protéines aient du mal à se placer dans les trous, ce qui nécessitera une augmentation de la concentration de l'échantillon. Même si de manière générale, la concentration optimale de l'échantillon dépend de la taille des particules, elle ne peut pas être rigoureusement identique pour l'ensemble des protéines (Vinothkumar and Henderson 2016) (figure 1.1.2.3-1).

Number of particles in projection/ $\mu\text{m}^2$  in 800 Å thick ice film (separation)

M.W.	Concentration				
	10mg/ml	2mg/ml	0.5mg/ml	0.1mg/ml	20 $\mu\text{g}$ /ml
10 kD	48000 (45Å)	10000 (100Å)	2500 (200Å)	500 (450 Å)	100 (1000 Å)
50 kD	10000 (100Å)	2000 (220Å)	500 (400Å)	100 (1000Å)	20 (0.2 $\mu\text{m}$ )
250kD	2000 (220Å)	400 (500 Å)	100 (1000 Å)	20 (0.2 $\mu\text{m}$ )	4 (0.5 $\mu\text{m}$ )
1 MD	500 (400Å)	100 (1000Å)	25 (0.2 $\mu\text{m}$ )	5 (0.4 $\mu\text{m}$ )	1 (1 $\mu\text{m}$ )
5 MD	100 (1000Å)	20 (0.2 $\mu\text{m}$ )	5 (0.4 $\mu\text{m}$ )	1 (1 $\mu\text{m}$ )	0.2 (2.2 $\mu\text{m}$ )
25 MD	20 (0.2 $\mu\text{m}$ )	4 (0.5 $\mu\text{m}$ )	1 (1 $\mu\text{m}$ )	0.2 (2.2 $\mu\text{m}$ )	0.04 (5 $\mu\text{m}$ )

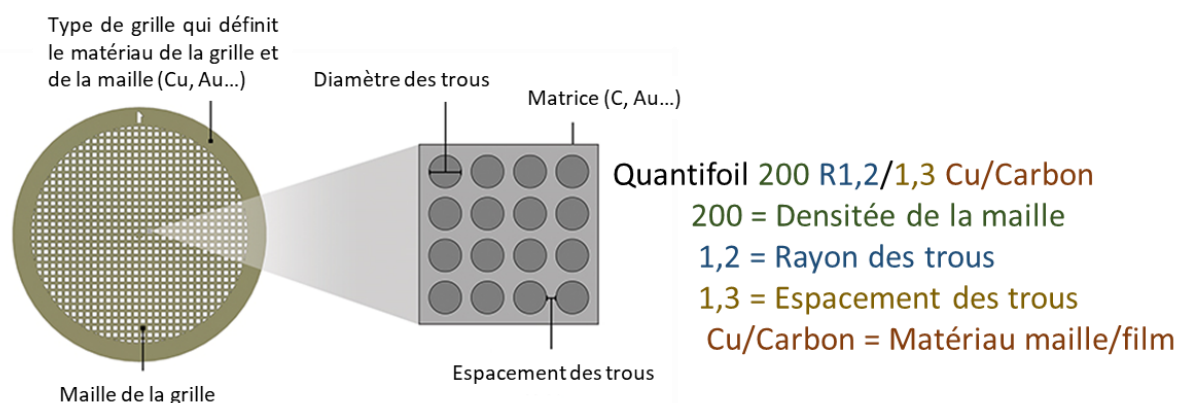
**Figure 1.1.2.3-1.** Figure reprise de Vinothkumar & Henderson, 2016, Cambridge University (Vinothkumar and Henderson 2016). Distribution attendue des particules sur des grilles Quantifoil. Le tableau donne le nombre attendu de particules uniques en cryo-EM par  $\mu\text{m}^2$  selon la taille en Dalton et la concentration en mg/ml. L'espacement attendu entre les particules est donné entre parenthèses. Les cases en rouge représentent une distribution trop dense, celles en bleu une distribution trop éparse et celles en vert une distribution acceptable. Il est impossible de donner des valeurs absolues car les particules se comportent très différemment en fonction du profil de la protéine. Par exemple, il se peut que toutes les particules collent au carbone, qu'elles fuient les trous, ou que la glace soit trop fine pour laisser les particules s'insérer.

Une fois vitrifiée, la grille est ensuite préparée pour le système de manipulation automatique de grilles du microscope (autoloader). Pour cela, la grille est déposée dans un anneau (ring) et fixée à celui-ci avec une agrafe en forme de C (C-clip). Jusqu'à 12 grilles peuvent être chargées en même temps dans un microscope de type Glacios/Krios à partir de la cassette de chargement, mais une seule peut être lue à la fois (figure 1.1.2.3-2).



**Figure 1.1.2.3-2.** Illustration des différentes méthodes de dépôt d'échantillon sur la grille (A) tirée de l'article de D'Imprima & Kühlbrandt 2021 (D'Imprima and Kühlbrandt 2021). Photos des différentes étapes de la préparation des grilles pour l'autoloader (clipping) (B). La photo en C montre une grille Ultrafoil fixée dans l'anneau (ring) par l'agrafe en forme de C (C-clip). Il est également possible d'observer la maille en or de la grille.

Enfin, pour pallier aux problèmes de distribution et de mauvaise densité de particules, il existe de nombreux types de familles de grilles. Elles se distinguent par le matériau qui les compose, généralement en cuivre ou en or et dont la matrice est en carbone ou en or. La grille est caractérisée par différentes spécificités telles que la taille des trous, l'espacement, ainsi que la densité de la maille (figure 1.1.2.3-3).



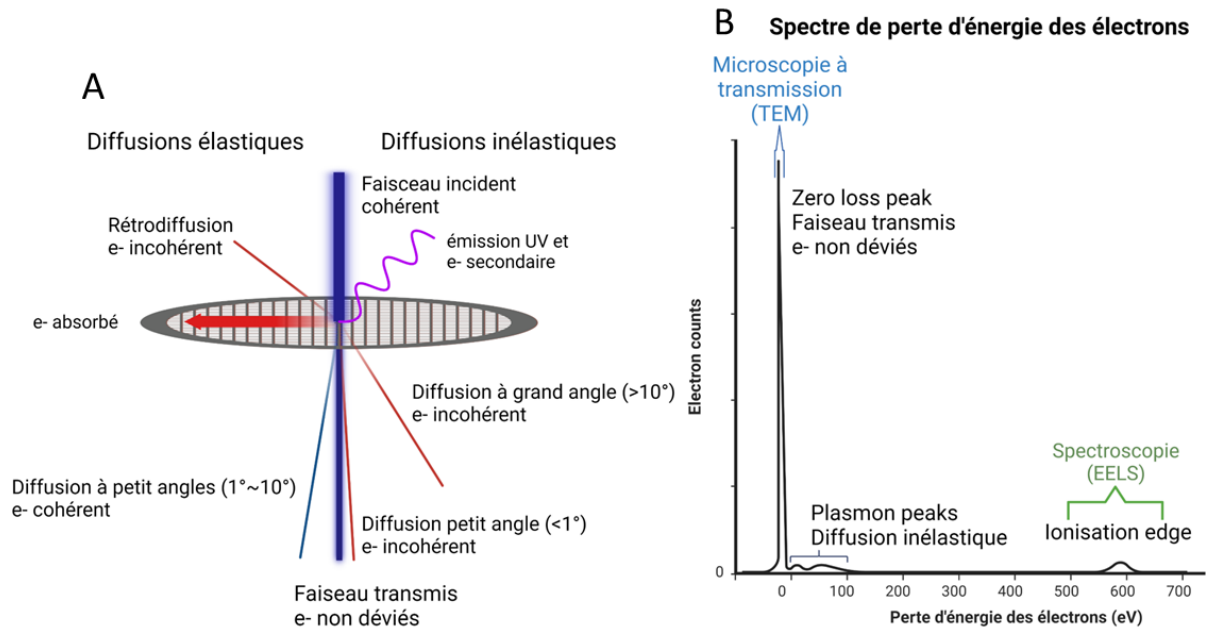
**Figure 1.1.2.3-3.** Image adaptée de "electron microscope science". Lexique de la grille de cryo-EM. En analyse SPA, les grilles utilisées sont principalement de la famille des Quantifoil. Ces grilles partagent une organisation commune qui consiste en une maille sur laquelle repose un support (carbone ou or) dans lequel des trous sont percés de diamètre et d'espacement régulier, ce qui facilite les collectes automatiques.

<b>Nom de la grille</b>	<b>Description</b>
<b>Quantifoil</b>	Grille en cuivre couverte d'un film de carbone de 10-12nm comme matrice, régulièrement percée de trous avec un diamètre contrôlé.
<b>UltrAufoil</b>	Grille en or couverte d'un film en or de 50nm comme matrice régulièrement percée de trous avec un diamètre contrôlé (Russo and Passmore 2016).
<b>UTC-Quantifoil</b>	Grille Quantifoil avec un fin film de graphène continu de 2-3nm qui recouvre la matrice percée (Naydenova et al. 2019).
<b>Lacey</b>	Grille avec un support en cuivre et une matrice en carbone avec des trous de formes et de tailles aléatoires.
<b>C-flat</b>	Grille Quantifoil avec des rebords de trous aplanis, en or ou en carbone.
<b>Graphène oxyde</b>	Quantifoil avec une couche de graphène hydrophile supplémentaire. Ces grilles n'ont en théorie pas besoin de traitement au plasma.

**Tableau 1.1.2.3-1.** Liste des principales grilles utilisées pour la résolution de la structure des protéines par cryo-EM SPA.

#### 1.1.1.4 Filtres à énergie et détecteurs

Lors de leur passage à travers l'échantillon, les électrons vont subir deux types de perturbations : les diffusions élastiques et les diffusions inélastiques (figure 1.1.2.4-1). En cryo-EM, il est possible d'améliorer la cohérence et donc le ratio signal sur bruit (contraste) de l'image avec un filtre d'énergie. Le filtre d'énergie fait passer les électrons dans un prisme magnétique qui va séparer les électrons en fonction de leur différent niveau d'énergie. Il est ainsi possible de stopper les électrons déviants à l'aide d'une fente centrée sur l'axe optique (energy slit). Les filtres d'énergie vont principalement retirer les diffusions inélastiques avec une résolution de déviation énergétique allant jusqu'à 10 eV pour les plus récents. Enfin, la perte d'énergie, bien que minime en TEM, est directement transmise à l'échantillon et le détruit ou génère un mouvement des particules qui réduit fortement la qualité de la reconstruction (D'Imprima and Kühlbrandt 2021). C'est pourquoi la dose maximale d'électrons qu'il est possible d'accumuler pour former l'image est relativement basse pour les échantillons biologiques. C'est l'une des grandes limitations de la cryo-EM.



**Figure 1.1.2.4-1.** (A) Représentation schématique des différentes diffusions qui ont lieu lors du passage du faisceau d'électron dans l'échantillon en cryo-EM. À gauche du faisceau, les diffusions élastiques ; à droite, les inélastiques. (B) Représentation schématique d'un spectre de perte d'énergie des électrons après avoir traversé l'échantillon, avec les régions indiquées. L'utilisation d'un filtre à énergie va retirer tous les électrons en dehors du pic "zero loss" en utilisant une fente (slit) centrée sur le pic et d'une résolution jusqu'à  $\Delta 10$  eV. Les régions de plus grande perte d'énergie (ionisation edge) sont utilisées dans les mesures de spectroscopie à perte d'énergie d'électron (EELS) qui permettent d'identifier les atomes qui composent le spécimen.

### 1.1.1.5 Cameras

Les cameras ont beaucoup évolué ces dernières années et sont responsables de la "resolution revolution" qui a fait exploser les performances de la cryo-EM (Kühlbrandt 2014; Vinothkumar and Henderson 2016; Hanske et al. 2018). Aujourd'hui les dernières caméras sont équipées de détecteur direct d'électrons capable de directement convertir la capture de l'électron en signal électrique, sans produire d'artefact. Les caméras les plus récentes produisent des films à très haute résolution de  $4096 \text{ pixels}^2$  de  $14\mu\text{m}$  et à très haute fréquences d'images par secondes (320 fps). Ces performances font que la caméra est rarement le facteur limitant de la résolution obtenue. Par exemple, pour une image collectée à un grossissement de  $\times 210000$ , avec une taille physique de pixel du détecteur de  $14\mu\text{m}^2$ , l'image obtenue aura un pixel de  $0,66 \text{ \AA}^2$ . Avec un tel pixel, la fréquence de Nyquist vaut  $1,22 \text{ \AA}$  ( $0,66 \times 2$ ), qui correspond à la résolution théorique de l'image. Or il est très rare d'atteindre de telle résolution en cryo-EM, qui est principalement limitée dans le niveau d'information par le faisceau lui-même et la protéine étudiée.

Durant ma thèse, j'ai pu voir émerger un nouveau format auquel sont enregistrées les images

en cryo-EM appelé "eer" (electron-event representation). Il s'agit d'une forme optimisée d'image qui permet d'alléger la taille des films de cryo-EM tout en apportant un plus haut niveau d'information dans l'image (Guo et al. 2020). Un jeu de données de cryo-EM est souvent composé de plusieurs milliers de films en 4K de 2 à 3 secondes enregistrés à très haute fréquence d'images (320fps) et pèse donc entre 2 et 3 TB. Il est donc normal de voir émerger un format optimisé spécifique pour la cryo-EM. Ce nouveau format amène une nouvelle façon de s'interroger sur la manière de paramétrer la collecte. En fonction du temps d'exposition (la durée du film) il est à présent nécessaire de noter combien de "sub-frames" ont été enregistrées en plus de l'information de la dose. Ces "sub-frames" seront regroupées par paquets pour former les "frames" lors du traitement des images. Il faudra alors choisir un diviseur qui permettra un tel regroupement, de façon que les frames cumulent une dose par surface d'environ  $1e^-/\text{Å}^2$ . Lors de ce groupement, le reste de la division (sub-frames en plus) ne sera pas utilisé par le logiciel de traitement et sera donc perdu. Il faut de ce fait optimiser la fraction en minimisant ce reste.

Prenons l'exemple pour un film collecté avec une dose par surface totale de  $60e^-/\text{Å}^2$  sur 756 sub-frames (2,36 secondes sur une Falcon 4i). Il est possible de le fractionner en 60 groupes, chaque groupe sera composé de 12 "sub-frames" et cumulera une dose de  $0,95e^-/\text{Å}^2$ . Mais, un tel groupement entraîne un reste de 36 sub-frames soit une perte de  $2,86e^-/\text{Å}^2$ . En revanche, en fractionnant les "sub-frames" en 63 groupes, toutes la dose du film sera utilisée, et chaque groupe contiendra toujours 12 "sub-frame" et donc la même dose par "frame". Il est aussi possible de réduire le nombre de groupe à 54 ce qui augmentera légèrement la dose de chaque groupe ( $1,1 e^-/\text{Å}^2$ ), améliorant ainsi la contraste de la "frame". S'éloigner de  $1 e^-/\text{Å}^2$  en augmentant trop la dose de la "frame" peut accumuler dans une seule "frame" trop de mouvement induit par le faisceau (beam-induced motion) (Brilot et al. 2012; Glaeser 2016) qui va flouter l'image de façon irréversible. À l'inverse, réduire trop la dose de la "frame" lui retire du contraste et peut se traduire par des mauvaises performances lors de l'alignement des "frames" du film.

Dans Relion il faut renseigner le nombre de "sub-frames" (que Relion appelle "frames") qu'il devra y avoir dans chaque frame ("fraction"). Alors que sur Cryosparc c'est le nombre de frames ("eer fraction") qu'il faut définir, dans lesquelles il va répartir également les sub-frames. Il en résulte un film avec n frames ( $n =$  "EER Number of Fractions" dans cryosparc) correspondant à la somme du même nombre de "sub-frames" initiales. Ces films sont ensuite convertis en micrographies par alignement et somme du signal des "frames".

#### 1.1.1.6 Plan et pilotage d'un microscope électronique à transmission

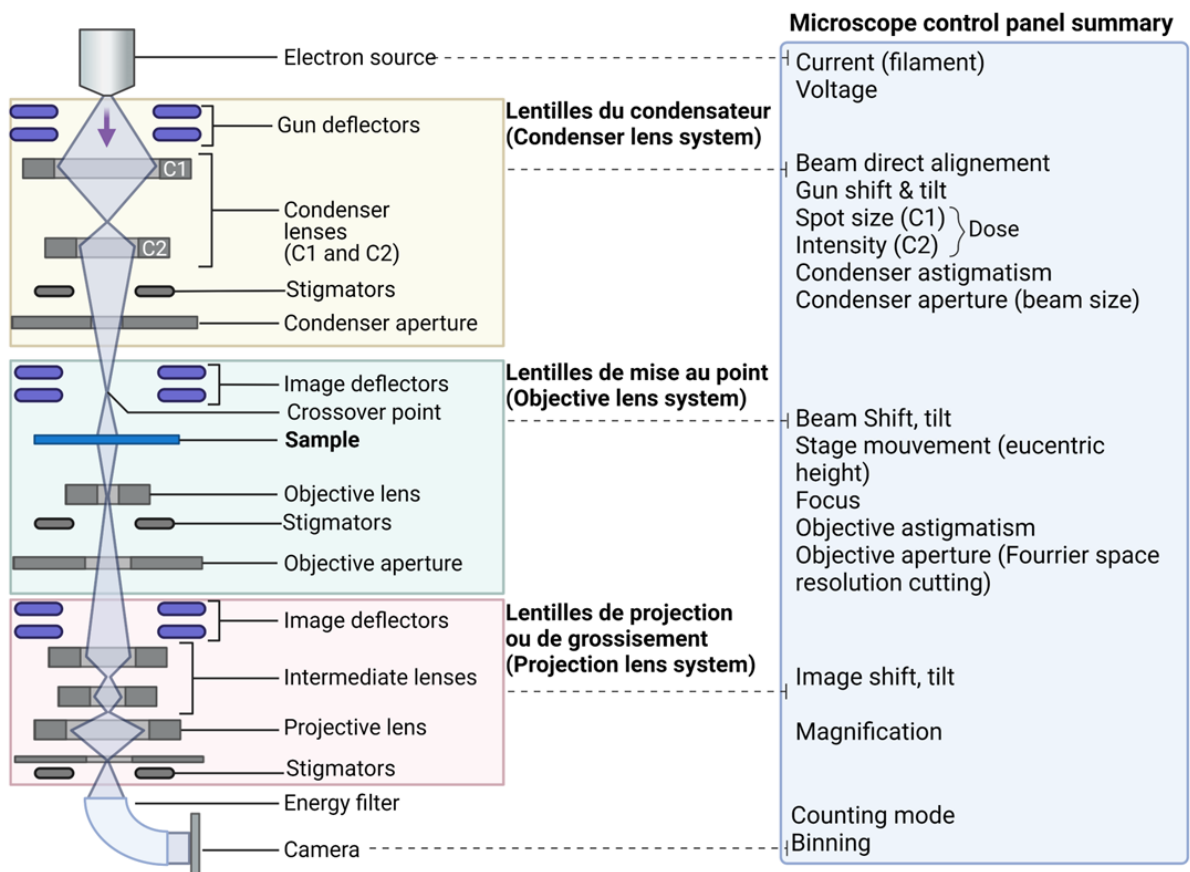
Un microscope est donc un assemblage le long d'une colonne des différents éléments présentés précédemment. L'intégralité de la colonne du microscope est dans un vite poussé. Elle est compartimentée en trois grandes parties : le système de lentilles condensatrices (condenser lens system), le système de lentilles de mise au point (objective lens système) et les lentilles de grossissement ou de projection de l'image (projection lens system) (figure

1.1.2.6-1).

Le système de condensateur ajuste le faisceau incident de l'échantillon. C'est ici que se font les réglages d'intensité, la parallélisation du faisceau, l'amélioration de sa cohérence notamment par le diaphragme (condenser aperture), et le choix de la taille du faisceau.

Sur le système de mise au point ou de focalisation il est possible d'agir sur l'angle et la direction du faisceau sur l'échantillon, la hauteur eucentrique, le point de pivot, et le diaphragme de l'objectif. A noter que ce diaphragme étant placé après la formation de l'image il va améliorer le contraste de l'image mais va réduire la résolution de celle-ci en agissant comme un filtre de fréquences dans l'espace de Fourier.

Enfin le dernier système contrôle le grossissement de l'image.



**Figure 1.1.2.6.** Liaison entre les systèmes du microscope et les panneaux de contrôle. Le microscope est divisé en trois systèmes de lentilles principaux : les lentilles du condensateur (jaunes), les lentilles de mise au point (vertes) et les lentilles de projection (rouges).



### 1.1.3 Traitement des images de l'analyse en particule unique (SPA) par cryo-EM

#### 1.1.1.7 Les trois principaux logiciels

Le logiciel **Cryosparc** (Punjani et al. 2017) est largement utilisé pour le traitement des images. Il a l'avantage d'avoir une interface graphique particulièrement bien optimisée et intuitive. Il est aussi communément reconnu comme le plus rapide dans les calculs. La communauté en ligne est très active, avec un forum largement fourni (<https://discuss.cryosparc.com/>). La principale contrainte est le format des fichiers (.cs) lisibles directement par Cryosparc mais pas par les autres programmes de cryo-EM et difficiles à manipuler. Il existe cependant des outils informatiques pour convertir vers le format ".star" beaucoup plus lisible (csparc2star de la suite pyem <https://github.com/asarnow/pyem>). Cryosparc est utilisable librement dans le milieu académique, mais nécessite une licence pour les laboratoires privés.

L'autre logiciel beaucoup utilisé est **Relion** (Scheres 2012), accessible gratuitement pour les utilisateurs académiques et du secteur privé. Il a été initialement développé par Sjors Scheres et bénéficie d'améliorations continues de la part d'une équipe diversifiée de chercheurs. Relion est réputé pour sa performance, sa richesse en termes d'outils et de personnalisation des tâches de calcul. La communauté en ligne est également très active (liste de diffusion CCPEM), tout comme la documentation (<https://relion.readthedocs.io/en/release-4.0/>). L'équipe derrière Relion se distingue par sa transparence concernant les algorithmes utilisés, ce qui est très apprécié. L'interface graphique peut sembler un peu moins intuitive, mais elle offre une facilité d'exécution en ligne de commande, similaire à celle de CCP4. Les calculs sont légèrement plus lents que ceux réalisés avec Cryosparc, et il peut y avoir quelques lacunes en ce qui concerne la visualisation et la rétroaction des résultats des calculs.

**Scipion** est apparu plus récemment (Rosa-Trevín et al. 2016) et tire sa force par sa capacité à prendre en charge les conversions de format des fichiers d'un logiciel à l'autre. Cela permet de jongler d'un algorithme à l'autre en toute simplicité et ainsi tirée avantage des forces spécifiques à chaque logiciel. Il propose également une palette d'outils de traitement et de diagnostic qui lui sont spécifiques. Il est pensé pour être facilement agrémenté par de nouveaux algorithmes développés par la communauté de chercheurs. Il est disponible gratuitement pour les industriels et les laboratoires public.

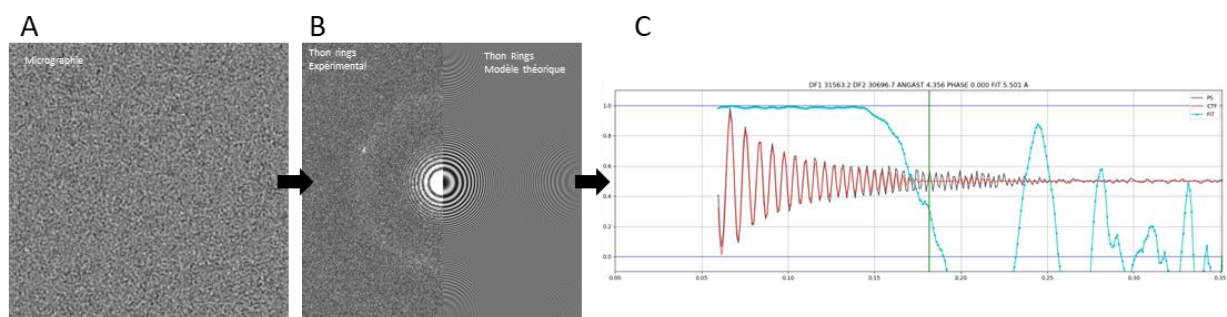
#### 1.1.1.8 Alignement des images et la fonction de contraste (CTF)

La première étape après une collecte est **d'aligner les "frames"** qui composent les films (motion correction) afin de moyenner le signal en prenant en considération les mouvements éventuels qui ont lieu pendant l'enregistrement du film. Cette étape crée les micrographies. Il est possible de corriger la dose totale du film en retirant les premiers et les dernières "frames", qui peuvent en plus souffrir de mouvement. Il est également possible de réduire la taille des micrographies pour gagner de l'espace et du temps de calcul.

La fonction de **transfert de contraste (CTF)** est l'expression de la qualité de l'information de l'image en microscopie électronique à transmission dans l'espace de Fourier. Elle exprime la contribution dans l'image de chaque fréquence spatiale, c'est-à-dire exprime les niveaux de résolution de l'image. Elle est la transformée de Fourier de la fonction d'étalement du point (PSF). Le concept mathématique et physique derrière la CTF repose sur les propriétés onde-corpuscule de l'électron. L'onde d'un électron unique, lorsqu'elle traverse la matière, sera diffusée dans toutes les directions mais avec la même longueur d'onde (car c'est un seul électron). En revanche, les ondes résultantes parcourent une distance différente entre le point de diffusion et le pixel qui les détecte. Elles ont donc un décalage de phase au moment de leur contribution cumulée sur le pixel et vont donc créer des interférences constructives ou destructives qui expliquent l'oscillation de la CTF. Pour chaque angle, il existe une onde résultante qui transporte un niveau d'information de plus ou moins haute résolution, c'est ce que l'on appelle la fréquence spatiale. Chaque onde transporte une composante de Fourier de l'image finale. Dans un système figé, certaines composantes sont donc perdues définitivement, mais il est possible de faire varier la manière dont les ondes interagissent les unes avec les autres en changeant la défocalisation. Enfin, un dernier concept important est la fonction d'enveloppe qui permet d'introduire les facteurs qui atténuent les contributions à haute fréquence spatiale (haute résolution). Tous ces concepts s'expriment dans le modèle mathématique de la CTF (CTF) et qui permettent au logiciel d'estimer correctement la CTF. Pour les échantillons biologiques, elle s'apparente à l'équation suivante (Zhang 2016) :

$$phase\ CTF(f) = -\sin\left(\pi\Delta z\lambda f^2 - \frac{1}{2}\pi C_s\lambda^3 f^4\right)$$

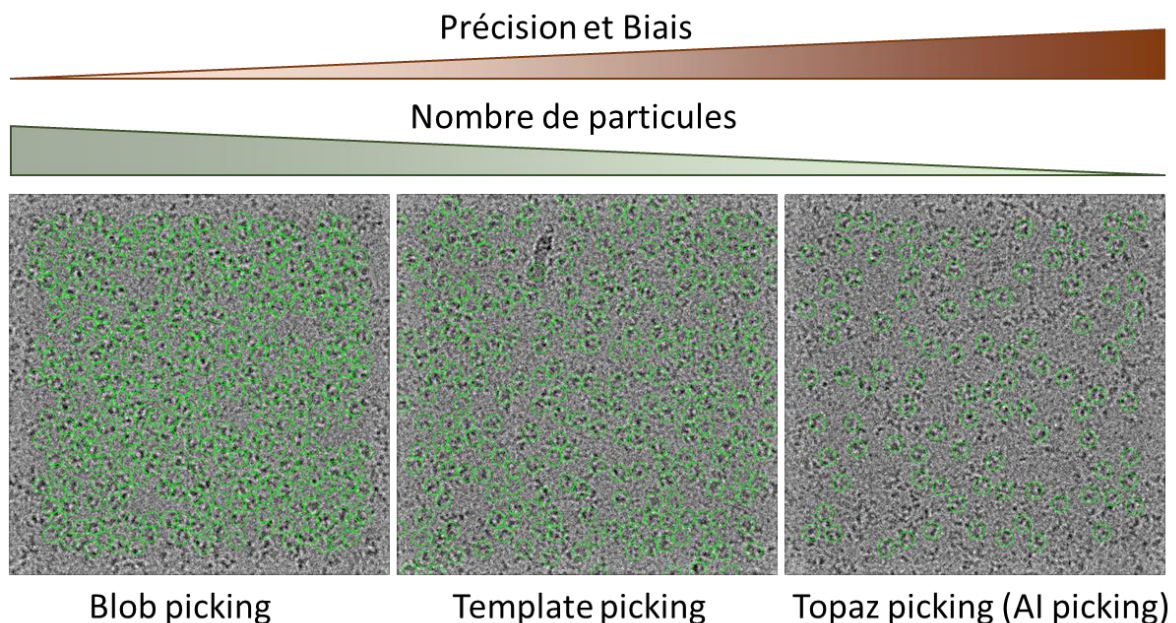
Avec  $\Delta z$  la défocalisation,  $\lambda$  est la longueur d'onde du faisceau d'électron,  $C_s$  est l'aberration sphérique, et  $f$  la fréquence spatiale. Tous les composants de l'équation sont constants au cours d'une collecte, excepté la défocalisation qui peut être changée d'une image à l'autre. Expérimentalement la CTF est retrouvée à partir du spectre de puissance de la micrographie (Thon rings) (Figure 1.1.3.2-1). Elle permet aussi d'estimer la résolution théorique, l'astigmatisme et la défocalisation de l'image.



**Figure 1.1.2.3.** Micrographie (A), Anneaux de Thon expérimentaux (gauche) et le modèle correspondant (droite) (B). Modèle de la CTF (courbe rouge) à partir du signal de Thon expérimental (courbe noire) et la corrélation entre les deux (courbe cyan) (C).

### 1.1.1.9 La recherche des particules

La recherche des particules sur les micrographies est un moment critique dans la reconstruction du volume. Il existe plusieurs approches possibles et souvent complémentaires. La première est le picking automatique basé sur les niveaux de contraste de l'image (picking Laplacian of Gaussian (Relion) & Blob picking (Cryosparc)), souvent utilisé en début de procédure pour extraire des motifs de l'image sans biais. La deuxième est la recherche à partir d'un modèle obtenu généralement par classification 2D (template picking). Enfin, il est possible d'extraire des particules via des algorithmes de "Deep Learning". (Topaz, SPHIRE-crYOLO).(Bepler et al. 2019; Wagner et al. 2019) (figure 1.1.3.3-1). Les coordonnées des particules sont ensuite utilisées pour découper une boîte autour et extraire l'image de la particule. Il est communément admis que la taille de la boîte doit correspondre à environ deux fois la taille de la particule qu'elle contient, pour contenir toute l'information délocalisée à cause de la défocalisation qui est surtout utile dans l'espace de Fourier.

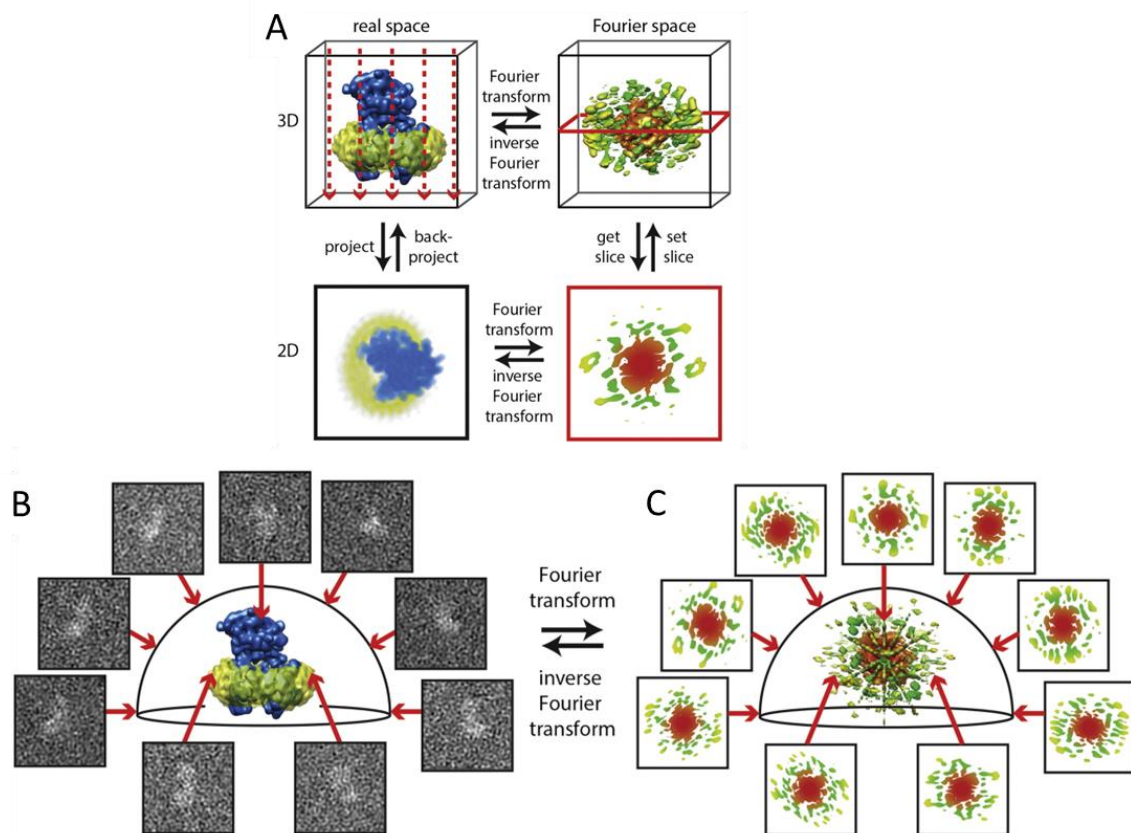


**Figure 1.1.3.3-3.** Illustration des différentes stratégies de picking. Blob à gauche, Template au centre et Topaz à droite. Chaque méthode a ses forces et ses faiblesses. Les approches par contraste (blob) ont peu de biais et trouvent plus de particules présumées mais sont peu précises. Les méthodes basées sur des modèles d'intelligences artificielles sont plus précises et génèrent moins de bruit mais sont biaisées. Ce biais peut être mis à profit pour extraire des orientations rares, lorsque les particules sont très disparates ou si la grille est de qualité moyenne ou basse à cause de contaminations.

### 1.1.1.10 La reconstruction et l'affinement des volumes

Le paradoxe en cryo-EM, c'est de travailler avec des images à très haute définition, mais sur lesquelles on ne voit pas grand-chose. L'astuce repose sur le moyennage du signal des images des particules pour atteindre la très haute résolution d'un volume 3D. Pour cela, les images

de l'objet, qui sont des projections de celui-ci selon différents angles, sont assemblées pour reconstruire le volume 3D dans l'espace de Fourier. Dans l'espace de Fourier, l'orientation des particules est représentée par des tranches de l'espace qui sont alignées les unes aux autres selon une ligne centrale continue, c'est le théorème de la tranche centrale (Nogales and Scheres 2015). Ensuite, le traitement consiste en des tours successifs de tri des particules selon différents critères, et d'affinement des alignements pour améliorer la reconstruction. Les plus grandes limitations dans la reconstruction sont : 1- le manque de contraste, provoqué par une glace trop épaisse ou un objet trop petit, 2- le manque de diversité dans les orientations, qui conduit à des effets anisotropiques, 3- ou une grande flexibilité qui limite le moyennage du signal.



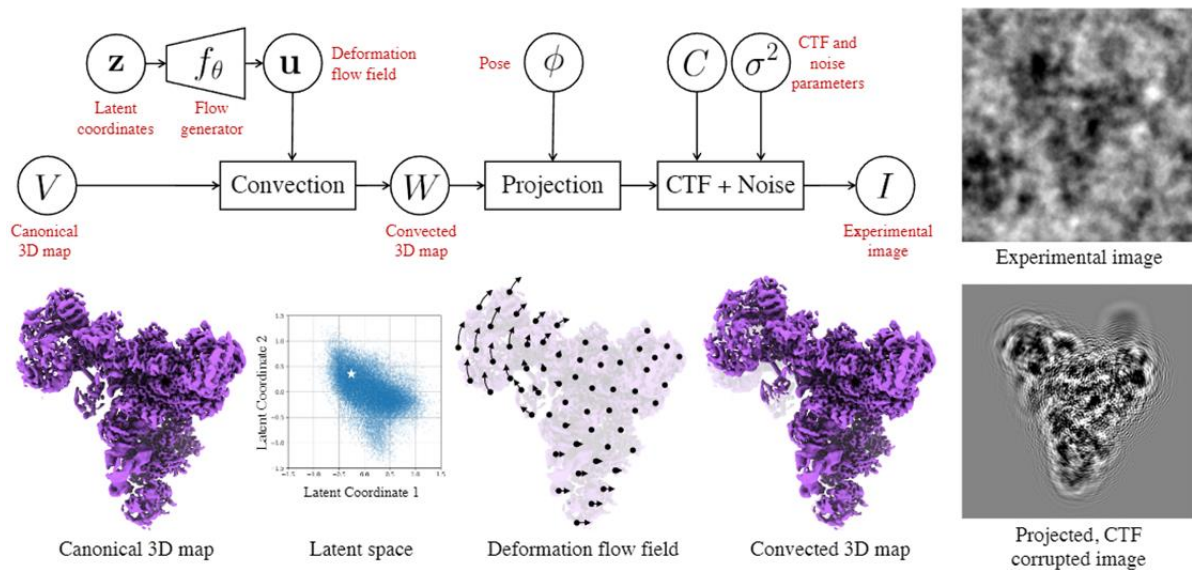
**Figure 1.1.3.4-1.** Figure reprise de Eva Nogales & Sjors Scheres 2015 (Nogales and Scheres 2015). Représentation schématique des étapes de la reconstruction du volume dans l'espace de Fourier. En **A** Le théorème de la tranche de projection. La projection 2D d'un objet 3D dans l'espace réel (gauche) est équivalente à l'extraction d'une tranche 2D centrale de la transformée de Fourier 3D de cet objet (droite). La direction de projection dans l'espace réel (à gauche, flèches rouges en pointillés) est perpendiculaire à la tranche (à droite, cadre rouge). En **B** les nombreuses projections expérimentales sont combinées et orientées pour créer une reconstruction 3D de l'objet dans l'espace réel en comparant les rétroprojections du volume avec les images expérimentales. L'alignement des particules est ensuite affiné dans l'espace de Fourier (**C**). Un processus itératif permet d'affiner les alignements en considérant de plus en plus de motifs haute résolution.

La résolution d'un volume est mesurée par la technique de la "Fourier shell correlation" introduite en 1987 par Marin Van Heel (Heel, 1987), puis améliorée plus récemment par Sjors Scheres et Shaoxia Chen en utilisant une approche dite "gold-standard". (Scheres and Chen 2012). Pour cela, les particules sont réparties aléatoirement en deux groupes distincts et indépendants, pour lesquels deux reconstructions 3D sont effectuées. À chaque itération, les deux volumes sont comparés dans l'espace de Fourier, et le score de cross-corrélation est tracé en fonction des fréquences spatiales. La résolution correspond à la fréquence pour laquelle le score tombe à 0,143 (Rosenthal and Henderson 2003). Il est possible de travailler sur les images des particules pour améliorer les données d'alignement et obtenir une carte de meilleure qualité. Par exemple le "Bayesian polishing" (Zivanov et al. 2019) va retravailler l'alignement des "frames" à l'échelle des particules. Ou bien une réestimation de la CTF par particule aide aussi à améliorer la carte.

Plus récemment, la tendance est aux algorithmes capables de considérer la flexibilité de la protéine. Les particules sont figées dans des conformations hétérogènes aléatoires qui peuvent être soit discrètes (enzyme, complexe protéique, protéines motrices...) soit continues (flexibilité). Parmi les algorithmes modélisant la flexibilité, on trouve 3DFlex récemment introduit dans Cryosparc (Punjani and Fleet 2021). 3DFlex est une approche par "deep-learning" pour extraire le mouvement continu de l'objet 3D à partir du signal des images expérimentales. C'est une approche itérative, où le modèle va, dans un premier temps, répartir les images des particules dans un espace latent.

Un espace latent est une méthode pour comparer la similitude entre des objets complexes de manière à ce que les objets les plus proches dans la dimension latente se ressemblent dans le monde réel. Cette méthode se concentre sur l'identification de schémas et des relations qui existent entre les variables. Elle est largement utilisée dans les approches basées sur le Deep Learning.

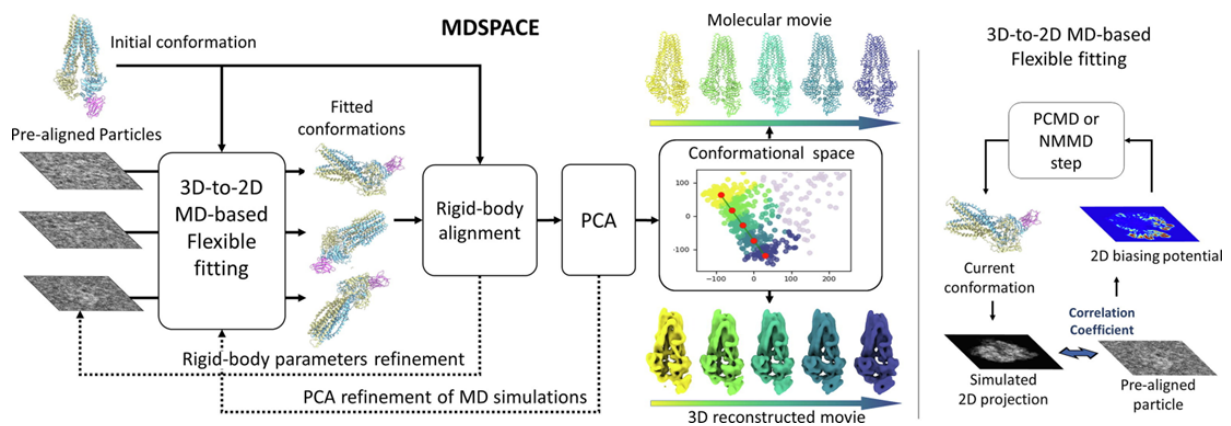
Le réseau neuronal va ensuite interpréter les mouvements et les retranscrire en déformant le volume à l'aide d'un champ de déformation. Les rétroprojections du volume déformé sont ensuite modifiées en fonction de la CTF, puis comparées aux particules expérimentales dans l'espace réel. En fonction des scores et au fil des itérations, l'espace latent et le champ de force sont optimisés pour identifier au mieux les mouvements. À la fin du processus, l'algorithme est capable de construire un volume consensus de meilleure qualité, notamment pour les domaines flexibles (Figure 1.1.4.3-2).



**Figure 1.1.2.4-2.** Image reprise de Punjani et al. 2021. Organigramme des différentes étapes de 3DFlex. 3DFlex représente la structure 3D flexible d'une protéine par la déformation d'une carte de densité 3D canonique ( $V$ ). Selon ce modèle, une image de particule unique est associée à une coordonnée latente de faible dimension ( $z$ ) qui code la conformation de la particule dans l'image. Un réseau neuronal ( $f_{\theta}$ ) convertit la coordonnée latente en un champ de déformation ( $u$ ) et un opérateur de convection déforme ensuite la densité canonique pour générer une carte déformée selon le mouvement identifié par le modèle ( $W$ ). Cette carte peut ensuite être rétroprojetée selon l'angle d'orientation  $\theta$ . L'image théorique est corrigée par CTF et comparée à l'image expérimentale. Le modèle s'affine au cours des itérations.

Un autre algorithme publié sur Scipion, nommé MDSPACE (Vuillemot et al. 2023), a une approche et un objectif différents. Cet algorithme est spécialement conçu pour identifier explicitement l'espace conformationnel de la protéine à l'échelle atomique. Cette approche a la particularité de mettre en relation des simulations de dynamique moléculaire avec les images expérimentales (Figure 1.1.2.4-3). MDSPACE s'inscrit dans la lignée des précédents algorithmes similaires, tels que HEMNMA & DeepHEMNMA (Harastani et al. 2020; Hamitouche and Jonic 2022). Tous ces algorithmes sont librement accessibles sur Scipion. Comme 3DFlex, MDSPACE utilise l'alignement préalablement attribué lors des affinements 3D, mais aussi un modèle atomique correspondant. Les modèles et les particules sont ensuite passés dans un bloc de simulation de dynamique moléculaire, au cours duquel les conformations sont triées par corrélation par rétroprojection aux images expérimentales. En même temps l'algorithme génère un potentiel de biais 2D qui guide les simulations suivantes. Ensuite, les modèles sont alignés de manière rigide (sans déformation de la chaîne principale) et analysés par composante principale (PCA). L'analyse en composante principale est une méthode qui permet de comparer des objets multidimensionnels dans une dimension réduite. Le PCA va ici directement représenter l'espace conformationnel de la protéine, auquel sont également associées les particules. Il est ainsi possible de reconstruire les volumes à partir des groupes identifiés dans l'espace conformationnel. La force de cet algorithme réside dans sa

capacité à lier directement les volumes au modèle atomique.



**Figure 1.1.2.4-3.** Image reprise de Vuillemot et al. 2023 (Vuillemot et al. 2023). Organigramme des grandes étapes de MDSpace. La simulation MD est guidée par un potentiel de polarisation 2D (à droite). Les lignes pointillées représentent le processus itératif, qui peut être répété plusieurs fois pour affiner l'espace conformationnel. NMMD = normal mode molecular dynamic, PCMD = principal component molecular dynamic.

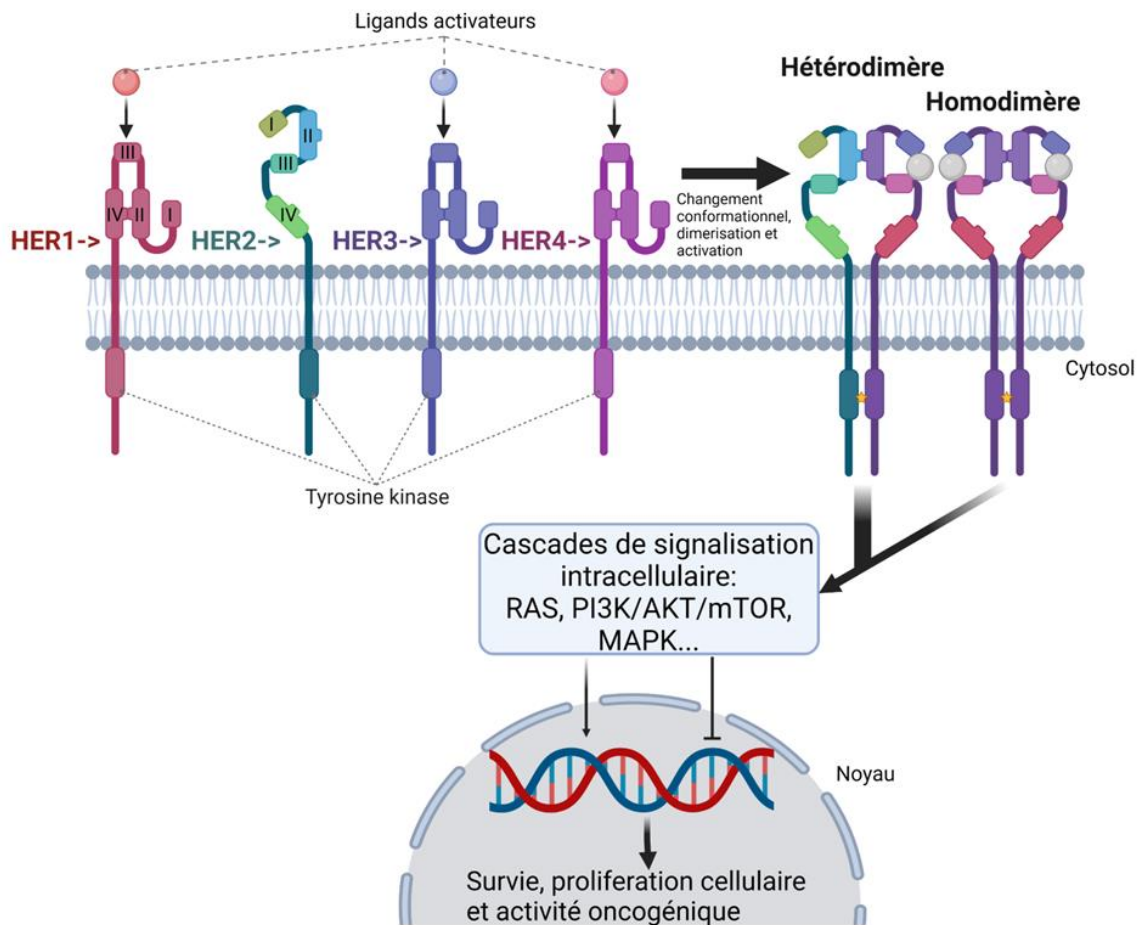
## 1.2 HER2-Trastuzumab-Pertuzumab

### 1.2.1 La famille des récepteurs transmembranaires ErbB

Les récepteurs du facteur de croissance épidermique humain (EGFR ou HER) sont des récepteurs transmembranaires appartenant à la famille ErbB (erythroblastic oncogene). Ils sont impliqués dans la régulation de la croissance cellulaire notamment par l'activité tyrosine kinase. Il existe quatre membres de la famille ErbB : EGF aussi appelé ErbB1 ou HER1, ErbB2/HER2, ErbB3/HER3 et ErbB4/HER4. Ils sont tous organisés en trois parties distinctes. Un domaine extracellulaire qui va médier la dimérisation des récepteurs et fixer différents ligands de croissance en fonction du récepteur, à l'exception de HER2 qui ne fixe aucun ligand. Ils ont ensuite un domaine transmembranaire et un domaine intracellulaire à l'activité kinase.

La formation d'homo et hétérodimère est médiée par la fixation d'un ligand par le domaine extra-cellulaire et entraîne un changement conformationnel de celui-ci. Ce changement va libérer le domaine II, rendant possible la dimérisation, qui se traduit par l'autophosphorylation du domaine intracellulaire (Diwanji et al. 2021). Le scénario est différent pour HER2 qui n'a aucun ligand connu et qui est constitutivement en conformation active (Figure 1.2.1-1) (Olayioye et al. 2000). Dans ce manuscrit nous traiterons de HER2, qui est le partenaire de dimérisation favori des autres récepteurs ErbB (Karunagaran et al. 1996). Les hétérodimères impliquant HER2 ont une plus forte capacité signalétique, dans le sens où ils induisent une prolifération et une survie cellulaire plus marquées (Karunagaran et al. 1996). La littérature fait également mention d'homo-dimérisation de HER2 basée sur des données de microscopie confocale et de STEM (Peckys et al. 2015; Peckys et al. 2021) ainsi que par des observations

in-vivo (Ghosh et al. 2011) mais sans donnée structurale.



**Figure 1.2.1-1.** En temps normal, *HER1*, *HER3*, *HER4* (mais pas *HER2*) sont maintenus en conformation inactive où le domaine extracellulaire est replié de façon que les domaines II et IV soient en interaction, empêchant toute dimérisation. Un changement de conformation intervient après la fixation d'un ligand au niveau des domaines I et III libérant le domaine II et permettant la dimérisation des récepteurs. La partie extra-cellulaire de *HER2* est constitutivement en conformation dépliée, ce qui contribue à en faire le partenaire privilégié des autres récepteurs. C'est la forme dimérique qui induit l'activation de diverses cascades signalétiques différentes selon les homo et hétérodimères formés. Ces signaux vont directement agir sur la régulation de gènes impliqués dans l'activation de la survie et de la prolifération cellulaire, et en cas de dérégulation, sont responsables d'une activité oncogénique.

Plusieurs facteurs sont à l'origine de la dérégulation de ces récepteurs, allant de l'augmentation à la réduction de leur expression, ou bien à des mutations du récepteur extracellulaire et de la kinase. Suite à ces perturbations, les récepteurs vont être systématiquement associés à divers cancers. EGFR (*HER1*) est majoritairement étudié pour



son implication dans les cancers pulmonaires autres que ceux "à petites cellules" (Non-Small-Cell Lung Cancer ou NSCLC) (Passaro et al. 2021). HER3 est un marqueur fort de certains cancers digestifs et gynécologiques (Jaiswal et al. 2013; Kojima et al. 2023). HER4 joue un rôle plus versatile en ayant parfois une activité antitumorale et d'autres fois au contraire favorisant une activité oncogénétique (Lucas et al. 2022). Il est tout de même impliqué dans les cancers colo-rectaux, des poumons, de l'estomac, le carcinome hépatocellulaire, des mélanomes ... (El-Gamal et al. 2021).

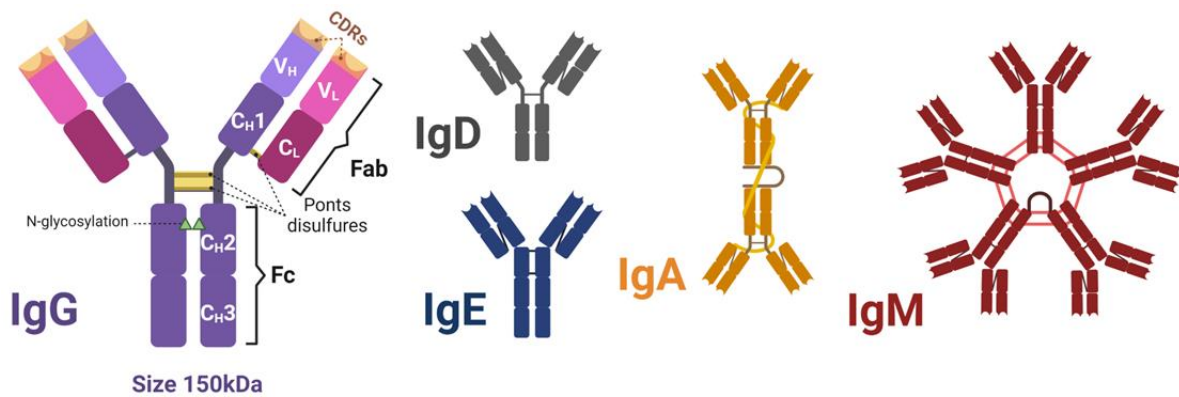
Enfin HER2 est aussi associé à de nombreux cancers, notamment dans une forme particulièrement agressive du cancer du sein dit HER2 positif (HER2+), où ce récepteur est surexprimé à la surface des cellules. Le cancer du sein est le premier cancer qui touche les femmes avec près de 30% des cancers dont 15 à 20% sont des cancers HER2+. HER2 est également impliqué dans 10 à 30% des adénocarcinomes gastriques. Dans les deux cas, le diagnostic en HER2+ est associé à un mauvais pronostic (Bouché and Penault-Llorca 2010; Fatemi et al. 2012) ce qui fait donc de HER2 une cible thérapeutique de choix.

Plusieurs stratégies ont été mises en place pour directement viser les récepteurs ErbB, telles que l'utilisation d'inhibiteurs de kinase (Geftinib, Lapatinib) ou l'utilisation d'anticorps monoclonaux (mABs) dirigés contre le domaine extra-cellulaire (Trastuzumab, Pertuzumab, Cetuximab, Panitumab) ainsi que d'autre dérivés d'anticorps (Bafisontama). Dans la suite de ce manuscrit nous allons traiter du Pertuzumab et du Trastuzumab.

### 1.2.2 L'origine et le rôle des anticorps

Les anticorps, aussi appelés immunoglobulines, jouent un rôle pilier dans l'immunité de tous les vertébrés. Ils ont été identifiés chez les mammifères, les oiseaux, les poissons et les reptiles avec des différences structurales, isotopiques et signalétiques mais ayant la même fonction principale de fixer une cible.

Un anticorps est une glycoprotéine d'origine naturelle en forme de "Y" issue de l'assemblage de deux chaînes lourdes (H) identiques et deux chaînes légères (L) identiques. La chaîne lourde est divisée en quatre domaines qui sont trois domaines constants " $C_H$ " et un domaine variable " $V_H$ ". Les domaines  $C_{H2}$  et  $C_{H3}$  forment le fragment cristallisable Fc, la région charnière et les deux ponts disulfure liant les chaînes lourdes entre elles lors de la dimérisation. Les domaines  $C_{H1}$  et  $V_H$  forment le domaine variable (qui interagit avec la cible) et la partie lourde du fragment Fab. La chaîne légère est composée d'un domaine constant  $C_L$  qui porte le pont disulfure avec la chaîne lourde et un domaine variable  $V_L$ . Elle complète la deuxième partie du domaine Fab. Le domaine Fab comporte trois régions déterminant la complémentarité (CDR) par chaîne. Ces CDRs sont responsables de la variabilité et de la spécificité de chaque anticorps. Enfin il existe différents sites de modification post traductionnelles (PTM) qui diffèrent selon l'isotype. Chez les IgG il y a une N-glycosylation sur le domaine  $C_{H2}$  (Figure 1.2.2-1). Chez les autres isotypes, il peut y avoir des N-glycosylations et des O-glycosylations à divers endroits de la chaîne lourde exclusivement.



**Figure 1.2.2-1.** Schéma des différents isotypes des immunoglobulines humaines. L'IgG est le principal anticorps dont l'architecture est à la base des autres isotopes IgA, IgD, IgE et IgM. L'IgG est formé par deux chaînes lourdes liées par deux ponts disulfures, et deux chaînes légères chacune liée à une chaîne lourde distincte. Une IgG est composée d'une région constante appelée fragment cristallisable Fc et une région variable de fixation à l'antigène Fab. La région du Fab comporte les régions déterminant la complémentarité CDR au bout du domaine V<sub>H</sub> responsable de la variabilité et de la spécificité des anticorps. Les autres isotypes sont des variantes de l'IgG.

Il existe cinq différents isotypes : IgA, IgD, IgE, IgG et IgM que l'on peut différencier principalement par leur structure, les voies signalétiques dans lesquelles ils sont engagés et leur localisation dans l'organisme. Les immunoglobulines IgG sont les plus représentées dans l'organisme et les plus concentrées dans la circulation sanguine. Les IgM sont des pentamères ou plus rarement des hexamères et sont les premières immunoglobulines secrétées par les lymphocytes B lors de la réaction immunitaire. Les IgA sont des dimères et sont retrouvées en grande quantité dans les muqueuses. Les IgE ont la particularité d'avoir un domaine C<sub>H4</sub> supplémentaire comme les IgM. Elles sont typiques des réactions allergiques. Enfin les IgD se distinguent par un temps de demi-vie court et n'ont pas une fonction clairement définie (Schroeder and Cavacini 2010).

L'immunité est un domaine complexe dont nous ne définirons que les fondamentaux nécessaires à la compréhension du manuscrit. Chez les mammifères, la réponse immunitaire se déclenche lorsqu'un pathogène ou une substance inflammatoire franchit la barrière épithéliale et entre dans l'organisme. La réaction qui en découle est scindée en deux parties qui sont l'immunité innée (non-spécifique) puis adaptative ou acquise (spécifique). La défense innée est dite non-spécifique dans le sens où elle est la première réaction de l'organisme à toute première exposition à un pathogène. Elle fait intervenir le système phagocytaire, les protéines de phase aiguë et du complément, les lymphocyte NK (Natural Killer), et les cellules dendritiques (Medina 2016). Les anticorps s'inscrivent dans le paysage de l'immunité plus tardive dite immunité acquise (ou adaptative). L'entrée dans cette phase est médiée par diverses voies signalétiques complexes et fait suite à l'invasion par un pathogène immunostimulant ou toutes substances immunogènes. Elle fait intervenir les lymphocytes B

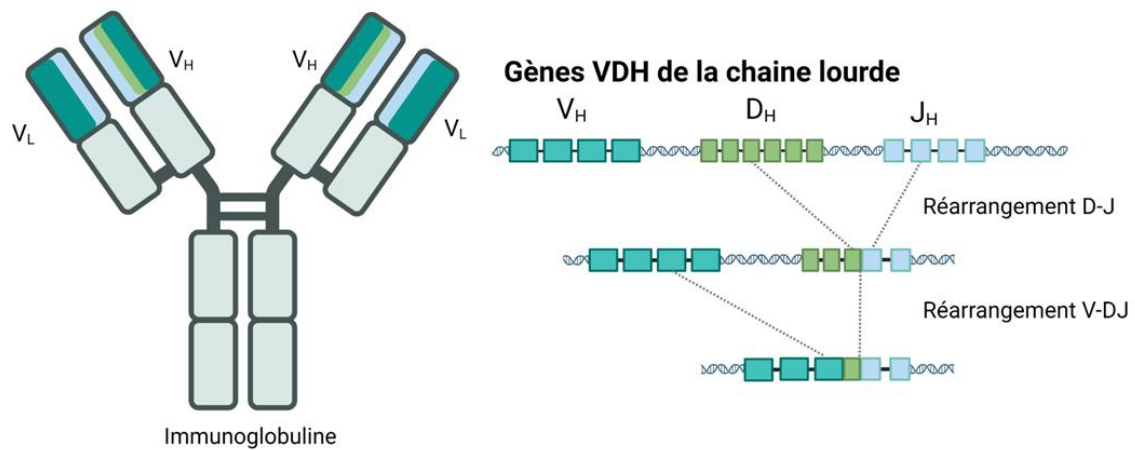
et T. Toutes la production d'immunoglobulines en circulation est assurée par les lymphocytes B tandis que les interactions ciblées entre les cellules sont assurées par les lymphocytes T. La particularité de cette voie réside dans l'attaque qui cible exclusivement l'agent à l'origine de son activation, à savoir l'antigène. Il existera donc une multitude de lymphocytes uniques, chacun capable de cibler spécifiquement un antigène. À noter que malgré le caractère monoclonal il peut y avoir une multi-spécificité où l'anticorps est capable de reconnaître plusieurs antigènes. Il résulte de cette réponse adaptative une mémoire immunitaire qui va persister et va permettre d'engager une réponse spécifique forte et immédiate lors de la prochaine rencontre avec la même substance immunogène / l'antigène.

### 1.2.3 L'origine de la diversité des anticorps monoclonaux.

L'une des grandes particularités de la production d'immunoglobuline est l'énorme diversité de la région se fixant à l'antigène (paratope), nécessaire pour répondre à toutes les agressions possibles. Cette diversité est rendue possible par l'unique processus de diversification des anticorps. Les gènes de la partie variable des lymphocytes sont organisés de manière atypique permettant le réarrangement somatique ou la recombinaison V(D)J. Cette recombinaison a lieu durant la maturation du lymphocyte, qui une fois mature produira des immunoglobulines ayant les spécificités de la séquence génétique unique.

Les gènes de la chaîne lourde sont organisés en trois clusters appelés Variable (VH), Diversité (D) et Joint (JH) qui vont être assemblés les uns aux autres d'une façon précise par un complexe d'enzymes appelées le "V(D)J recombinase" (Roth 2014). Le nombre exact de gènes est variable d'un individu à l'autre mais nous pouvons tout de même estimer la diversité combinatoire. Le cluster VH est composé d'environ 60 gènes, le cluster D d'environ 27 gènes et le JH de 6, chacun de ces gènes vont se réarranger de manière aléatoire pour former un enchaînement VDJ. Il y a donc une combinaison d'environ 9700 possibilités.

Du côté de la chaîne légère, une première variabilité est apportée par la formation de deux familles de chaînes, kappa ( $\kappa$ ) ou lambda ( $\lambda$ ). Si les gènes viennent du chromosome 2, la chaîne légère sera une chaîne kappa ( $\kappa$ ), s'ils sont du chromosome 22 ce sera une chaîne lambda ( $\lambda$ ). Le chromosome 2 contient 40 gènes  $V\kappa$  et 5 gènes  $J\kappa$  soit environ 200 régions variables légères kappa différentes. Le chromosome 22 contient 30 gènes  $V\lambda$  et 4 gènes  $J\lambda$  soit près de 120 combinaisons uniques possibles. Les chaînes lourdes et légères s'assemblent pour former un anticorps complet. L'anticorps provenant d'un tel assemblage est dit monoclonal, car issu d'une cellule lymphatique B unique (Janeway et al. 2001; Mayer and Hudrisier 2012; Little et al. 2015). Ce processus de diversification combinatoire amène à une première diversification d'au moins  $3,1 \times 10^6$  combinaisons possibles (Figure 1.2.3-1).



**Figure 1.2.3-1.** Illustration adaptée de la revue "molecular biology of B cells – (second edition) 2015" (Little et al. 2015). Schéma de la recombinaison génétique V(D)J pour la chaîne lourde. Les couleurs des gènes à droite renvoient à la région variable correspondantes à gauche.

En réalité cet exercice mène à une estimation très éloignée de la réalité. Durant le processus il se produit des perturbations aléatoires qui font exploser la diversité génétique (Rose 1982). La première est la diversification jonctionnelle, provoquée par de mauvaises recombinaisons qui décalent les cadres de lectures. La seconde est l'hypermutation somatique qui se produit durant la phase de mémoire immunitaire. Ce processus introduit un grand nombre de mutations lors de la division des lymphocytes. Cette phase est surtout à l'origine de la maturation d'affinité. Ces deux éléments amènent à estimer la diversité à au moins  $10^{12}$  immunoglobulines uniques, soit le nombre de grains de sable de 0,3 mm de diamètre (soit un volume  $\sim 10^{-11} \text{ m}^3$ ) nécessaire pour remplir un réservoir de  $10 \text{ m}^3$ , l'équivalent de 1,7 réservoirs de Canadair. Certaines études font même état de  $10^{16}$  à  $10^{18}$  (Briney et al. 2019; Rees 2020) soit l'équivalent du nombre de grains de sables occupant un volume de 10 millions de mètres cubes soit  $1/10^{\text{ième}}$  de la dune du Pilat ou 4000 piscines olympiques !

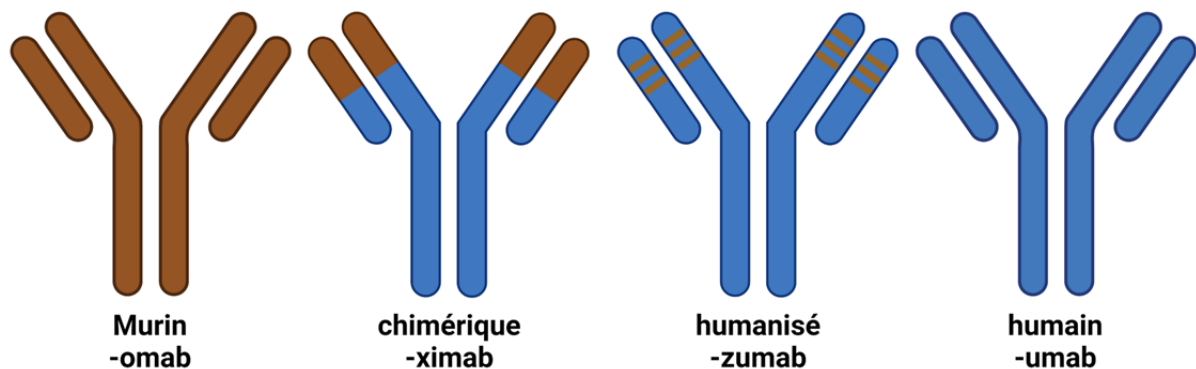
L'exercice nous montre à quel point la nature a rendu possible la création d'un nombre démesurément grand de paratopes uniques et ainsi la mise à disposition d'un arsenal varié d'IgG capables de cibler toutes les protéines. C'est une propriété qui est au cœur des stratégies thérapeutiques mais qui peut aussi être détournée pour diverses autres applications, dont certaines seront illustrées au cours de ce manuscrit.

#### 1.2.4 Les méthodes de productions et l'humanisation des anticorps.

La première production de mAb à large échelle a été faite à partir de cellules murines en 1975 par le biochimiste César Milstein et l'immunologiste George Köhler, qui se rencontrent au cours d'un séminaire à l'université de Basel. Ils mettent en place une méthode (KÖHLER and MILSTEIN 1975) qui permet d'immortaliser des lymphocytes provenant de souris immunisées via un antigène d'intérêt, en les fusionnant avec des cellules de myélome. La cellule ainsi immunisée s'appelle un hybridome. À partir de ces hybridomes, un clone peut être isolé et mis en culture pour ainsi exprimer un anticorps issu d'un unique clone dirigé contre un épitope

spécifique. Ces travaux ont donné lieu à un prix Nobel de physiologie et médecine en 1984 et sont encore aujourd'hui à la base de la production d'anticorps monoclonaux à application thérapeutique et diagnostique.

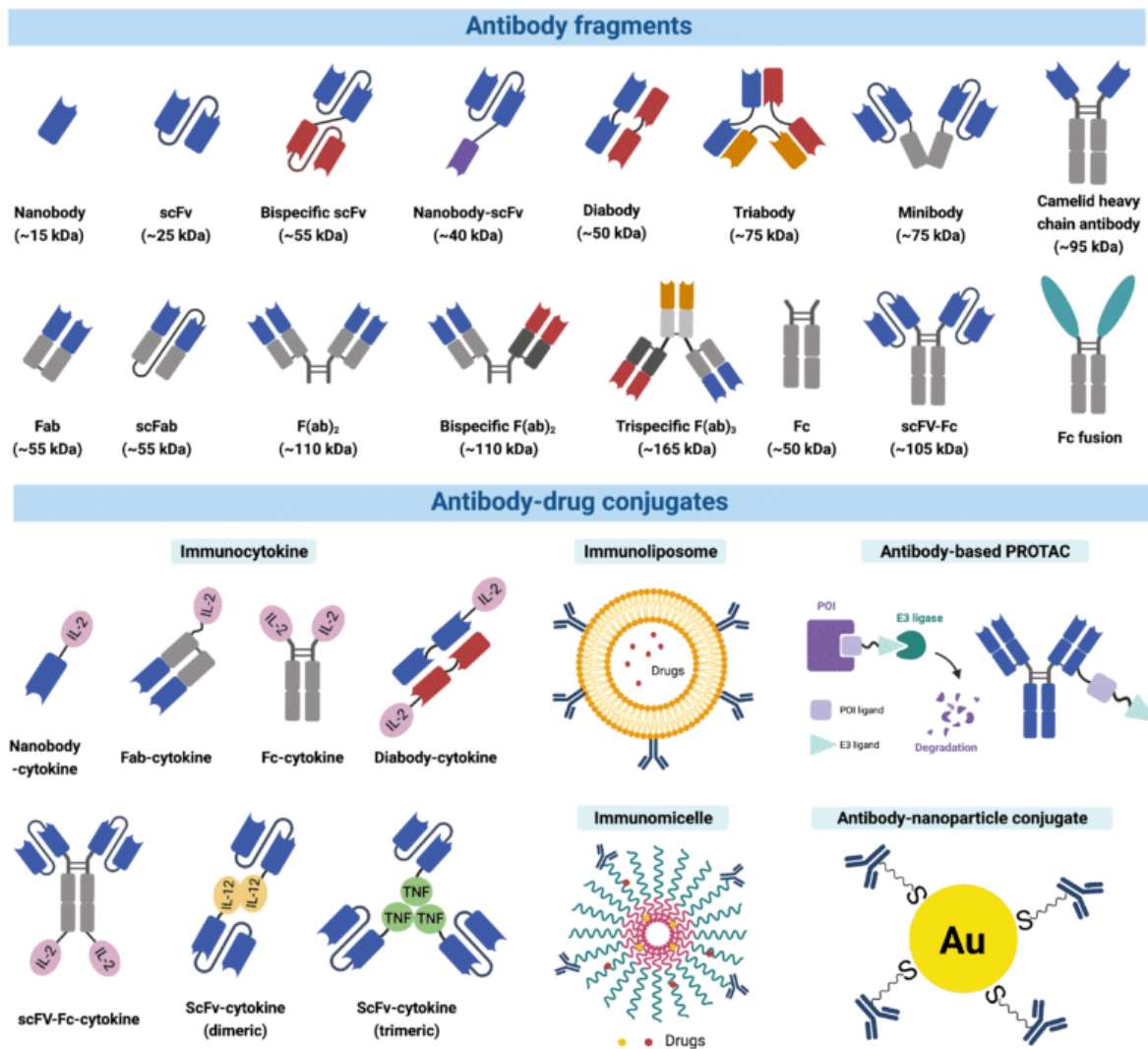
Les anticorps peuvent aussi être produits par la culture de cellules eucaryotes de type HEK ou CHO, cette approche est moins chère et plus rapide que l'approche classique par hybridome. Pour cela il faut se procurer la séquence de l'anticorps, le plus communément par une approche de type display (Ledsgaard et al. 2018). Celui-ci pourra donc aussi être travaillé et modifié en ingénierie génétique. Parmi ces modifications, on trouve l'humanisation des mABs qui consiste à rendre la séquence de l'IgG non-humain plus proche de celle d'un anticorps humain. Ceci permet d'améliorer notamment son efficacité thérapeutique et son acceptation par l'organisme (Safdari et al. 2013). Leur nomenclature est normalisée, ce qui permet de connaître simplement leur niveau d'humanisation. Les noms avec le suffixe -omab correspondent à des anticorps exclusivement murins, ayant surtout une application en biochimie et en diagnostic. Le suffixe -ximab est spécifique aux anticorps dit chimériques. Seuls les domaines  $V_H$  et  $V_L$  sont encore murins et les anticorps sont donc humains à ~67%. Les anticorps au suffixe -zumab sont dit humanisés et seules les parties hypervariables (CDRs) sont encore murines ce qui en fait un anticorps à 90% humain. Enfin le suffixe -umab désigne les mABs entièrement humains (Figure 1.2.4-1).



**Figure 1.2.4-1.** Les différentes familles d'anticorps humanisées. Les anticorps monoclonaux sont nommés selon le niveau de séquence humaine.

### 1.2.5 Les dérivés d'anticorps

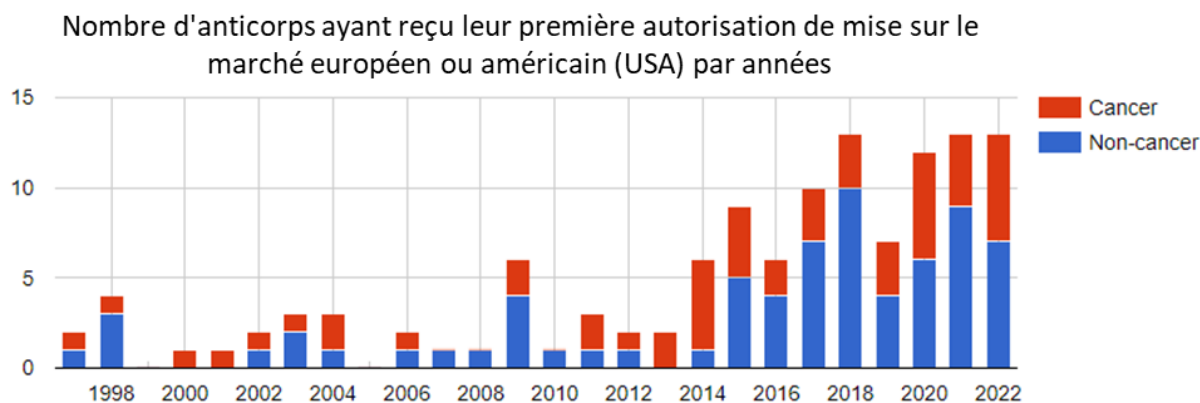
Il existe de nombreux types de variants d'anticorps naturels (Fabs, les Nanobody, anticorps de camélidés...) ou issus d'ingénierie protéique (ScFv, anticorps bi ou trispécifique, anticorps conjugués ...) (figure 1.2.5-1). Cette diversité conduit à une application large tant du point de vue médical qu'en recherche.



**Figure 1.2.5-1.** Illustration de chez RapidNovor créée par le Dr Yuning Wand disponible sur <https://www.rapidnovor.com/what-are-monoclonal-antibodies/>. Les dérivés d'anticorps. En haut les dérivés de types fragments. En bas les différentes variantes d'anticorps conjugués.

## 1.2.6 Les applications des anticorps

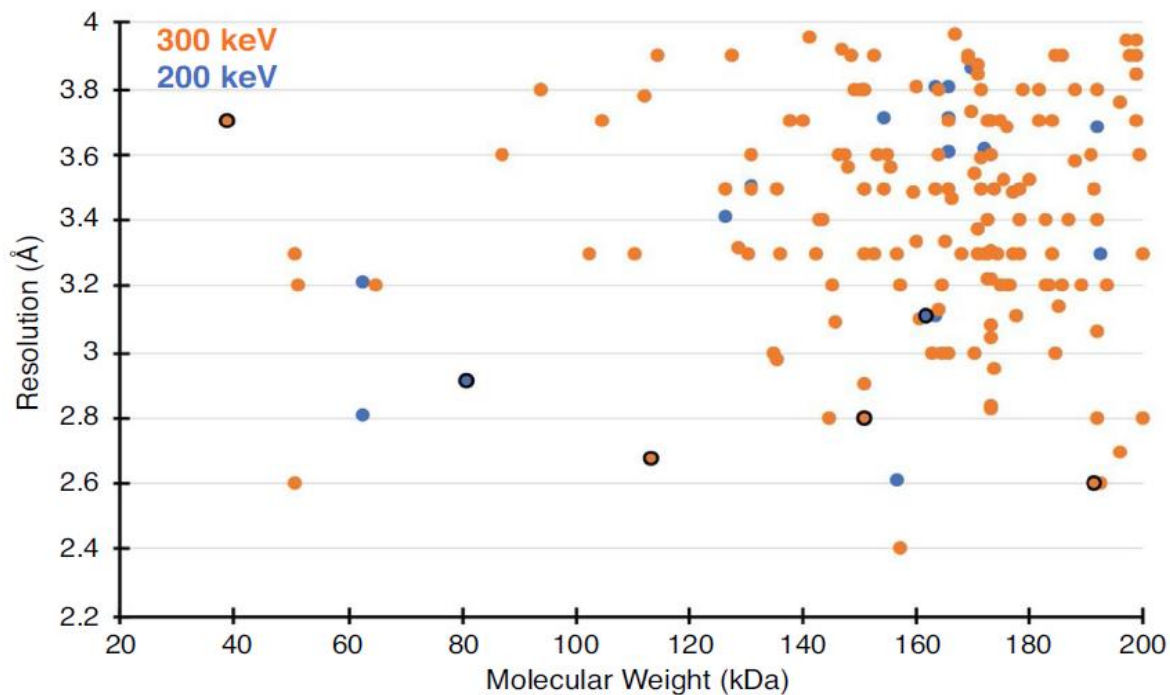
Dans ce manuscrit nous allons illustrer deux applications typiques des mABs. La première est l'application thérapeutique. Ces dernières années nous observons une forte croissance du nombre d'anticorps disponibles sur le marché du médicament (figure 1.2.6-1). Cette tendance s'explique par l'optimisation des méthodes de sélection et de production des mABs d'une part et de leurs propriétés intrinsèques qui les lient au système immunitaire.



**Figure 1.2.6-1.** Histogramme provenant de la société des anticorps (<https://www.antibodysociety.org/resources/approved-antibodies/>) (AntibodySociety 2023) montrant le nombre d'anticorps ayant reçu une autorisation de mise sur le marché chaque année depuis 1998. Données uniquement pour l'europe et les USA. En rouge les anticorps et dérivés en application anticancéreuse, et en bleu pour toute autre application. Les dernières années ont vu le nombre d'autorisations exploser.

L'autre application qui s'inscrit dans ce manuscrit concerne plutôt les dérivés d'anticorps, utilisés par exemple en vectorisation de médicaments (Figure 1.2.5-1), mais qui peuvent aussi être très utiles pour augmenter la taille d'un objet d'intérêt. En effet en Cryo-EM il y a une gamme de taille en dessous de laquelle il est difficile d'avoir une bonne carte de cryo-EM (Wu & Lander, 2020) (Figure 1.2.6-2). Ainsi, il est possible d'utiliser les fragments de mAB pour artificiellement grossir une cible et l'étudier par Cryo-EM. De plus, un fragment est bien moins flexible qu'un mAB et va ainsi améliorer le signal et donc la carte. L'utilisation de Fab par exemple permet de gagner 50kDa par Fab fixé. Les nano-bodies et scFv sont aussi très utiles car encore moins flexibles qu'un Fab tout en augmentant significativement la taille des objets de 15kDa et 25kDa respectivement. Il existe une multitude d'autre applications des anticorps et leurs dérivés comme en microscopie confocale, en biochimie, en cristallogénèse etc....

Cependant les mABs coûtent chers à produire et il faut généralement compter dans les 2700€ pour 1 mg d'anticorps. Le rendement de conversion en Fab est souvent de l'ordre de 30 à 40 %, l'utilisation d'une telle approche peut rapidement s'avérer coûteuse est peu rentable. Nous allons voir qu'il existe des alternatives au anticorps bien plus adapté pour l'aide à l'analyse structurale.



**Figure 1.2.6-2.** Graphique tiré de Mengyu Wu & Gabriel C Lander 2020 (Wu and Lander 2020). Exemple du nombre de structures résolues par cryo-EM classé en fonction de leur taille et de la résolution de la carte.

### 1.2.7 Trastuzumab

Le Trastuzumab est un anticorps monoclonal humanisé de type IgG1. Il se fixe sur le domaine IV de la partie extracellulaire de HER2, ci-après désignée sHER2. La première approbation pour un usage thérapeutique contre le cancer du sein HER2+ est en 1998 par la Food and Drug Administration (FDA) aux USA (FDA 1998) suivie par l'Agence européenne des médicaments en 2000 (EMA 2000). Il est également utilisé aujourd'hui contre certains cancers gastriques métastatiques HER2+ (HAS 2016).

L'activité anticancéreuse du Trastuzumab s'effectue par un cumul d'effets biologiques notamment un effet cytostatique, une régulation angiogénique, ainsi que l'agrégation et l'internalisation des récepteurs HER2 (Austin et al. 2004; Ben-Kasus et al. 2009). L'internalisation est d'autant plus importante que l'anticorps est utilisé avec un autre anticorps ciblant un épitope différent (Ren et al. 2012; Leyton 2020). Enfin, il existe aussi un effet inhibiteur de clivage protéolytique, empêchant la formation d'une forme alternative de HER2, nommée p59HER2, constitutivement active (Molina et al. 2001). D'un autre côté il apporte aussi les effets associés au domaine Fc, notamment la cytotoxicité anticorps dépendante (ADCC), permettant une interaction directe avec le système immunitaire. Enfin le Fc est aussi le responsable de la longue demi-vie de l'anticorps améliorant ses propriétés pharmacocinétiques.

Cependant il est vite apparu que l'utilisation du trastuzumab seul résulte en l'apparition de



résistances qui se traduisent par une perte de réponse des patients au traitement par le trastuzumab. Parmi les mécanismes de résistance sont principalement avancés le masquage de l'épitope, l'augmentation de la cascade signalétique de HER2, la compensation par des voies signalétiques alternatives, et enfin l'échec du recrutement du système immunitaire (Pohlmann et al. 2009).

### 1.2.8 Pertuzumab

Le Pertuzumab est aussi un anticorps humanisé de type IgG1 ciblant HER2 développé pour une utilisation en combinaison avec le Trastuzumab. Il a reçu une première autorisation de mise sur le marché par la FDA en 2012 (FDA 2012) suivi l'année suivante par l'EMA (EMA 2013). La synergie médicamenteuse est rapidement apparue lors de l'utilisation combinée de cet anticorps avec le Trastuzumab (Nahta et al. 2004; Richard et al. 2016; Nami et al. 2018). Le Pertuzumab se fixe sur le domaine II de sHER2. Le premier effet de sa fixation est le blocage stérique de toute dimérisation de HER2 (Nami et al. 2018) en plus des effets bénéfiques associés au Fc. Il est également certainement acteur de l'internalisation de HER2 lorsqu'il est utilisé avec le Trastuzumab et est aujourd'hui couramment utilisé dans les immunothérapies dans le traitement du cancer du sein (Scheuer et al. 2009; Blumenthal et al. 2013; Howie et al. 2019; Liu et al. 2022).

### 1.2.9 Bilan des structures disponibles

À la suite de leur mise sur le marché, plusieurs travaux d'études structurales par cristallographie et par microscopie électronique ont apporté la description atomique de la fixation entre les mABs et HER2. Toutes les structures ont été réalisées à partir du Fab et non de l'anticorps entier, le Fab étant moins flexible que l'anticorps entier. Plusieurs structures haute résolution ont été obtenues pour des complexes partiels (Cho et al. 2003; Franklin et al. 2004; Diwanji et al. 2021). En revanche il n'existait qu'une seule structure du complexe ternaire, obtenue en 2019 à moyenne résolution (Hao et al. 2019) qui nous a servi de référence pour réaliser les travaux de thèse. Nous verrons dans ce manuscrit comment nous avons réussi à publier en 2023 la première structure haute résolution du complexe ternaire permettant de décrire pour la première fois avec précision les deux épitopes en simultanément avec une résolution locale proche des 2,5Å.

Code PDB	Description	Technique	Res. (Å)	Date	Publication et auteurs
1s78	HER2/Pertuz	X-ray	3,25	2004	Insights into ErbB signaling from the structure of the ErbB2-pertuzumab complex (Franklin et al. 2004)
1n8z	HER2/Tastuz	X-ray	2,52	2003	Structure of the Extracellular Region of HER2 Alone and in Complex with the Herceptin Fab (Cho et al. 2003)
7mn5-8	HER2/HER3/ NRG1 $\beta$ Trastuz	Cryo-EM	3,45	2021	Structures of the HER2-HER3-NRG1 $\beta$ complex reveal a dynamic dimer interface. (Diwanji et al. 2021)
8Q6J	HER2/Trastuz /Pertuz	Cryo-EM	3,2Å	2023	Atomic structure and conformational variability of the HER2-trastuzumab-pertuzumab complex (Ruedas et al. 2023)
6oge	HER2/Trastuz /Pertuz	Cryo-EM	4,36	2019	Cryo-EM structure of Her2 extracellular domain-Trastuzumab Fab-Pertuzumab Fab complex (Hao et al. 2019)
6bgt	HER2/Trastuz mutant	X-ray	2,7	2017	To be published ( <i>Christie, M., Christ, D. not published</i> )
3be1	HER2/Trastuz (dual)	X-ray	2,9	2009	Dual specific bH1 Fab in complex with the extracellular domain of HER2/ErbB-2 (Bostrom et al. 2009)

**Tableau 1.2.9-1.** Bilan des différentes structures disponibles des complexes HER2-Pertuzumab, HER2-Trastuzumab et HER2-Pertuzumab-Trastuzumab. Les couleurs indiquent les meilleures résolutions pour les complexes dimériques (gris et bleu) ou trimériques (orange).

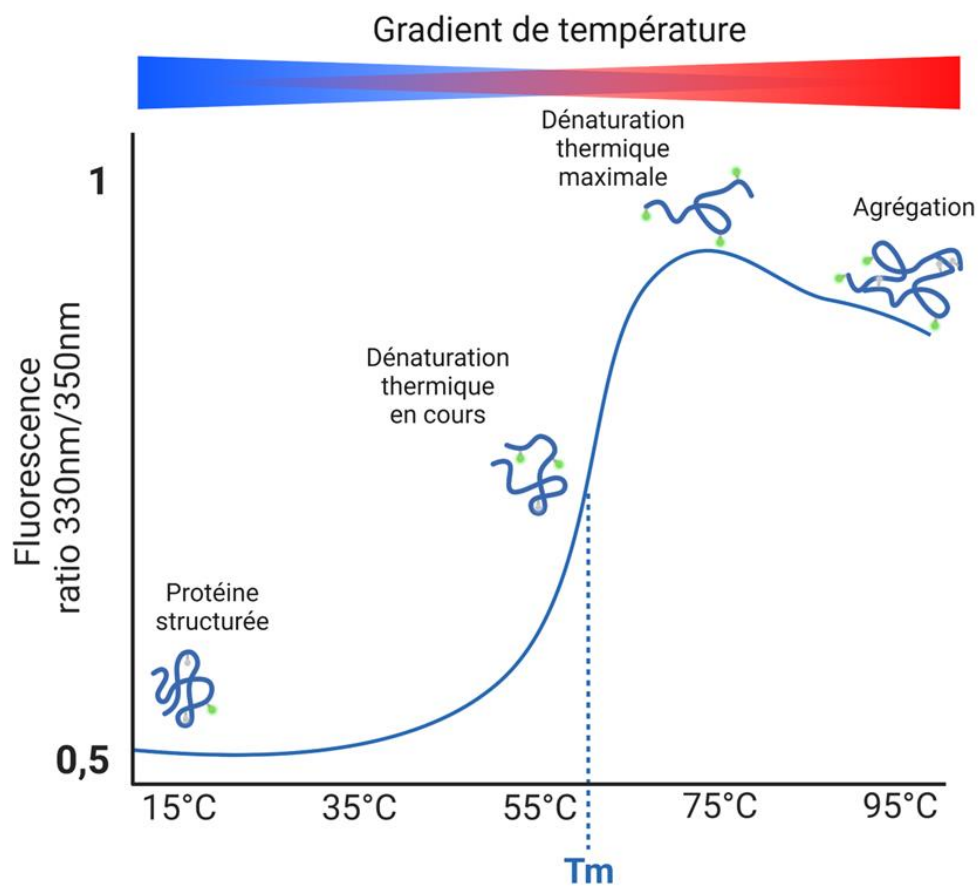
## 1.3 Principes biophysiques

### 1.3.1 Nano-DSF

La fluorimétrie différentielle à balayage micro-échelle (nano-DSF) permet d'obtenir des informations sur la stabilité de la protéine via la température de fusion ( $T_m$ ). C'est une méthode en solution, sans marquage mais destructrice. L'instrument excite l'échantillon à 280 nm et analyse le signal de fluorescence des tryptophanes à 330 nm et 350 nm. La protéine en solution est soumise à un gradient de température de 15 °C à 95 °C qui va entraîner une dénaturation thermique. Cette dénaturation se traduit par un changement de l'exposition des tryptophanes et donc une variation de la fluorescence. C'est souvent le ratio 350/330 qui est tracé ce qui permet d'accentuer la transition de signal. Le  $T_m$  correspond à la température à mi-hauteur de la variation de fluorescence (Figure 1.3.1-1). La mesure est faite sur un Prometheus Panta de chez Nanotemper, également capable de mesurer la diffusion de la

lumière et la "backreflexion" liée à l'agrégation. Le  $T_m$  est utilisé pour rendre compte de l'impact positif ou délétère de l'environnement de la protéine sur sa stabilité. La fixation d'un ligand ou le changement du tampon sont souvent associés à un changement du  $T_m$  favorable ou défavorable. Cette technique s'inscrit parfaitement dans les stratégies de criblage haut débit de ligands et d'optimisation de tampon.

Cette technique nécessite environ 10  $\mu$ l d'échantillon. La concentration minimale est variable en fonction du coefficient d'extinction molaire de la protéine et du nombre de tryptophanes participant au signal.



**Figure 1.3.1-1.** Schéma simplifié des évènements mesurables par nano-DSF. Le concept de base repose sur l'excitation à 280 nm et la mesure de la réponse des tryptophanes à 330 nm et 350 nm tout au long d'un gradient de température de 15 à 95 °C. La dénaturation thermique est associée à un changement du signal d'absorbance d'où est extrait la température de dénaturation ou de fusion ( $T_m$ ). La décroissance de fluorescence parfois observée en fin de gradient est généralement provoqué par l'agrégation des protéines. Des scénarios à plusieurs paliers sont aussi observables dans le cas où différents composants sont présents en solution.

### 1.3.2 DLS

La mesure de diffusion dynamique de la lumière (DLS) est également réalisée en solution sans marquage et elle est parfois non destructrice selon l'instrument utilisé. La mesure repose sur le principe de la diffusion de Rayleigh selon lequel le faisceau de lumière monochromatique (405 nm sur le Prometheus Panta) va subir une diffusion élastique au contact de petites particules telles que des protéines. La DLS mesure la corrélation de la fluctuation de l'intensité du faisceau au cours du temps. Il existe une relation entre le mouvement Brownien et la taille de la particule (Einstein 1905), qui est directement associée à la variation du signal enregistré, et à partir de laquelle il est possible d'extraire le rayon hydrodynamique par l'équation de Stokes-Einstein :

$$Rh = \frac{kT}{6\pi\eta D}$$

*Rh est le rayon hydrodynamique, k et la constante de Boltzmann, T est la température en Kelvin, η est la viscosité du solvant et D est le coefficient de diffusion de la protéine.*

La mesure expérimentale de la fluctuation permet d'obtenir la fonction d'autocorrélation de second ordre noté  $g^2(q;\tau)$  directement dépendante du vecteur d'onde (q) et du temps de corrélation (τ). À partir de l'égalité de Siegert  $g^2(q;\tau)=f[g^1(q;\tau)]$ , il est possible d'extraire la valeur du taux de décroissance (Γ) par l'équation  $g^1(q;\tau)=\exp(-\Gamma \tau)$ . Enfin le taux de décroissance permet de retrouver le coefficient de diffusion transverse (D) par l'équation :

$$\Gamma = Dq^2 = D\left(\frac{4\pi n}{\lambda} \sin \frac{\theta}{2}\right)^2$$

*q est le vecteur d'onde, n est l'indice de réfraction, λ la longueur d'onde du laser, et θ l'angle de diffusion.*

Du fait de la polydispersité de l'échantillon,  $g^1(q; \tau)$  ne peut être exprimé par une exponentielle simple et se complexifie en une série d'intégrale en fonction de Γ. Il existe différentes régressions mathématiques pour analyser les données. La plus commune est la méthode du "Cumulant" qui considère l'échantillon comme une unique particule et représente la distribution comme une distribution gaussienne unique (Scotti et al. 2015). L'indice de polydispersité est la variance relative de la distribution gaussienne soit pour un pic unique :

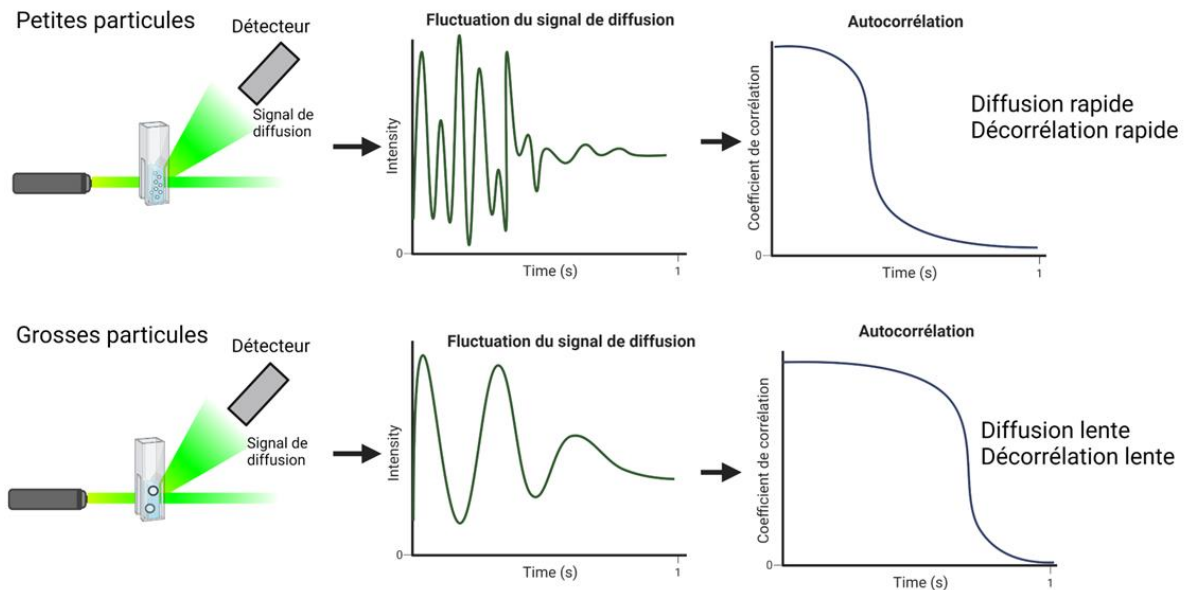
$$PdI = \frac{\sigma^2}{\bar{X}^2}$$

*avec σ = l'écart type et  $\bar{X}$  = taille moyenne de la gaussienne*

Les autres méthodes comme "Contin" ou "Corenn" vont considérer l'échantillon comme multimodal. Ces dernières ont pour objectif de retranscrire au mieux le spectre de distribution des tailles de l'échantillon. Contin est une approche par l'équation de l'intégrale inverse (Provencher 1982). Corenn est un algorithme à base d'intelligence artificielle utilisé par

certains fabricants (LS Instruments).

La DLS peut être mesurée soit en capillaire (NanoTemper Panta) (ce sera une mesure destructive dans le sens où l'échantillon sera perdu), soit dans une cuve en quartz (Malvern Panalytical) après quoi l'échantillon est récupéré. Les capillaires consomment jusqu'à 10 µl, la cuve à besoin de 30 µl. La concentration minimale dépendra de l'échantillon.



**Figure 1.3.2-1.** Illustration du principe de la DLS. La fluctuation du signal de diffusion élastique du faisceau monochromatique est directement liée au mouvement Brownien, lui-même lié à la taille de la particule. La mesure de la corrélation du signal au cours du temps permet l'extraction du rayon hydrodynamique de la particule en solution.

### 1.3.3 SEC-SLS

Le principe de la diffusion statique de la lumière (SLS) est très proche de celui de la DLS. Elle est réalisée à 660 nm. La SLS utilise le même principe que la DLS mais mesure la moyenne de l'intensité du signal. À partir de cette mesure il est possible de déterminer avec précision la masse de la protéine en solution à partir de l'équation de Debye-Zimm. La taille des protéines étant généralement inférieure à  $\lambda/20$  l'équation s'exprime ainsi :

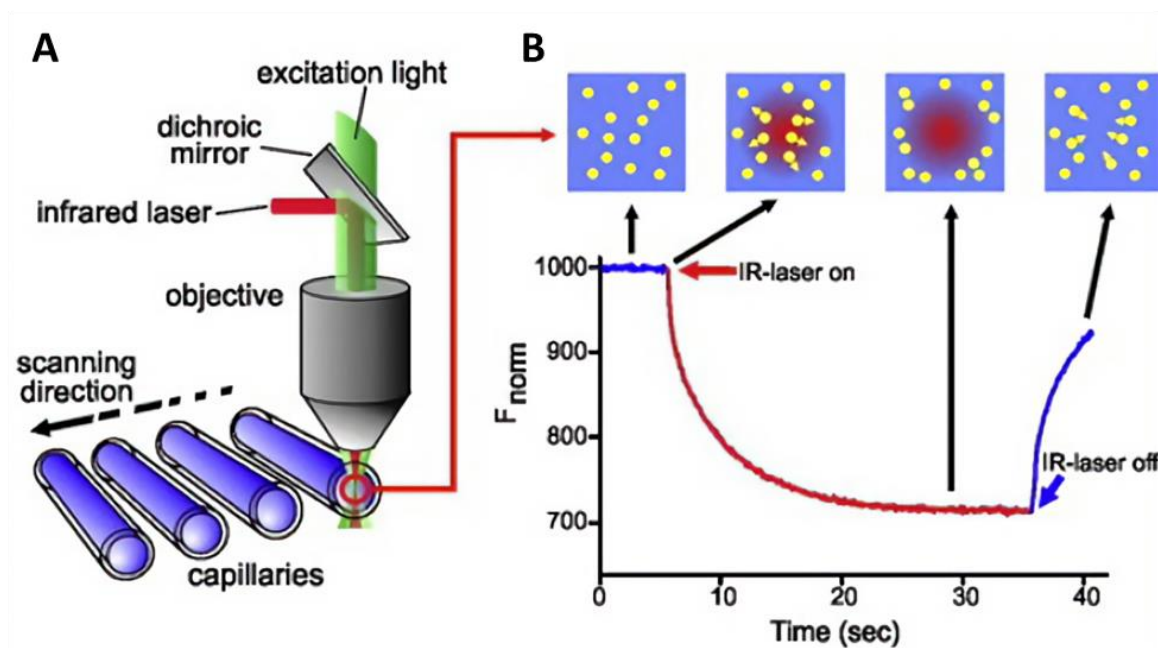
$$\frac{KC}{\Delta R} = \frac{1}{Mw} + 2A_2C$$

$K$  est la constante optique liée à l'instrument,  $A_2$  est le coefficient de Viriel lié aux propriétés du fluide,  $\Delta R$  est le ratio de Rayleigh ou l'écart d'intensité avec la lumière incidente,  $C$  est la concentration de l'échantillon et  $Mw$  est le poids moléculaire de la protéine.

### 1.3.4 Micro-Scale Thermophoresis

La thermophorèse micro-échelle (MST) est une méthode pour quantifier les interactions biomoléculaires. Elle consiste à suivre l'évolution de l'intensité de la fluorescence d'une zone à la suite de l'application d'un gradient de température via un laser infra-rouge. C'est une mesure en solution et destructive. La fluorescence de la protéine, marquée ou non, est mesurée à température ambiante, un laser infra-rouge est appliqué localement et va augmenter le mouvement brownien. Les molécules vont se déplacer ce qui va faire varier la fluorescence locale. Le laser IR est ensuite stoppé et l'instrument mesure le retour à l'équilibre (Figure 1.3.4-1). La variation de la fluorescence dépend de la taille, la charge, la couche d'hydratation et la conformation de la protéine. Ces propriétés sont altérées par la fixation d'un ligand ou d'un partenaire. Ainsi en réalisant plusieurs mesures à différentes concentrations il est possible d'extraire l'affinité entre le ligand/partenaire et la cible.

Cette mesure présente l'intérêt de consommer peu d'échantillon, autour des 160  $\mu\text{l}$  à  $\sim 50$  nM pour effectuer les mesures complètes permettant l'extraction du  $K_d$ . La présence de détergent n'a pas d'impact majeur sur la mesure (il peut perturber la fluorescence du marqueur mais cela ne sera pas limitant).

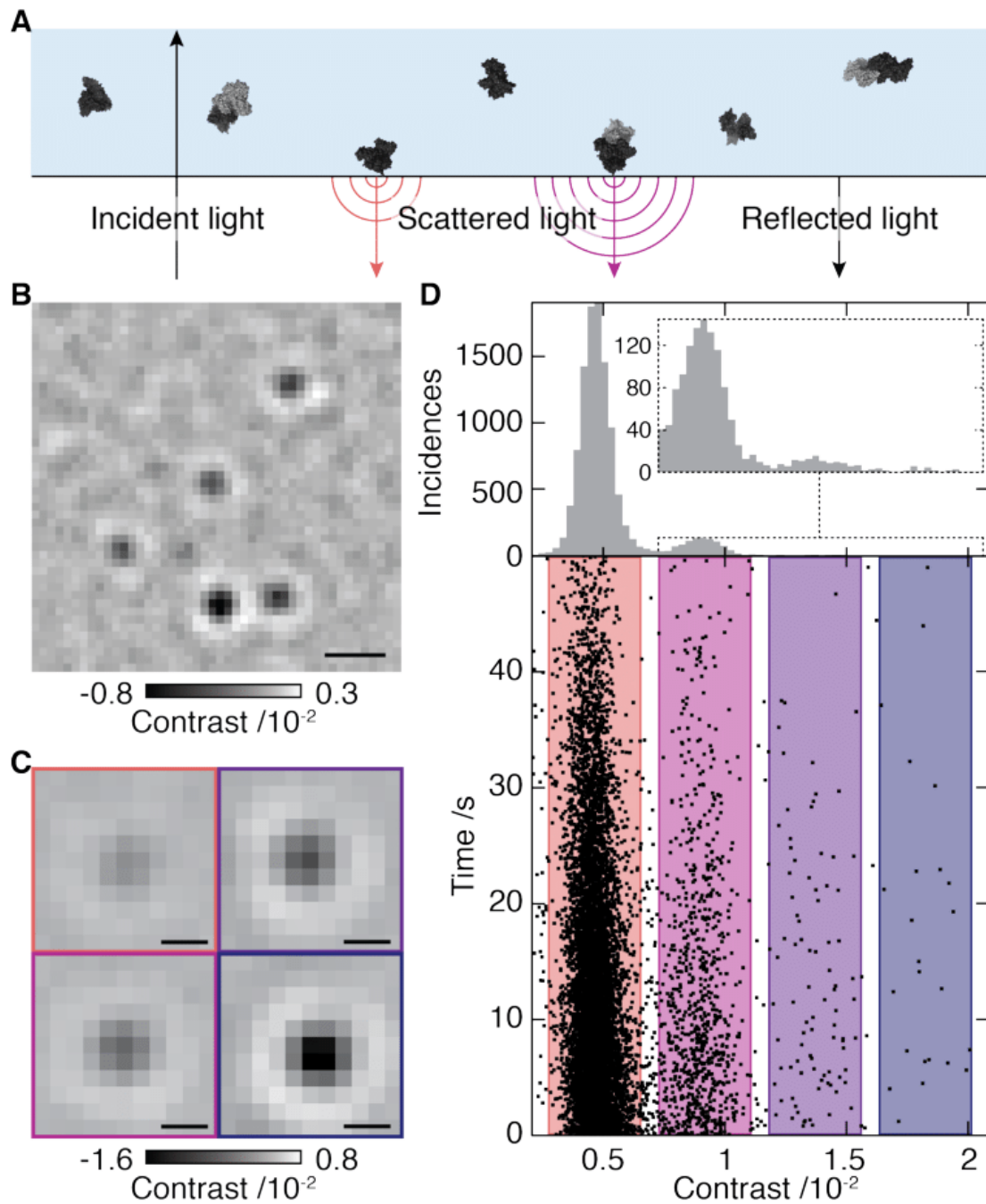


**Figure 1.3.4-1.** Image reprise de Jerabeck-Willemsen et al., 2014 (Jerabek-Willemsen et al. 2014) . Illustration du principe de la MST. En **A**, schéma du système optique, avec le miroir dichroïque qui permet de dévier le laser infra-rouge vers le capillaire tout en permettant l'excitation et la captation de la fluorescence par le même objectif. À droite (**B**) le schéma du mouvement des particules en solution à la suite de l'application du gradient de température. En dessous le type de graphique typique qui est obtenu par MST au cours des différents évènements.

### 1.3.5 Photométrie de masse

La photométrie de masse, ou spectrométrie de masse par diffusion interférométrique, est une technique relativement récente qui permet de mettre en relation la taille de la particule avec l'intensité de la diffusion de la lumière lors de son contact avec une surface en verre (Young et al. 2018). C'est une mesure en solution sans marquage. Contrairement à la DLS ou la SLS, elle mesure chaque évènement de façon ponctuelle, ce qui permet donc l'attribution des populations même avec un échantillon fortement hétérogène (Figure 1.3.5-1).

Afin d'assurer la détection d'évènements uniques, cette technique nécessite de se placer à faible concentration, de l'ordre de la dizaine de nanomolaire. C'est à la fois une force et une faiblesse dans le sens où elle nécessite très peu de matériel (10  $\mu$ l) mais s'éloigne beaucoup des conditions biochimiques standards de la biologie structurale. De ce fait, certains complexes supporteront mal cette dilution qui peut entraîner une dissociation. De même la présence de détergent dans l'échantillon perturbe le signal et peut être limitant. Celui-ci ne pouvant généralement pas être dilué librement, il peut masquer les signaux de la protéine ou provoquer des faux positifs. La gamme d'application est relativement large de 30kDa à 5MDa avec une résolution d'environ  $\pm 20$  kDa en fonction du niveau de disparité de chaque population.



**Figure 1.3.5-1.** Image reprise de Young G. et al., 2018 (Young et al. 2018). A chaque contact d'une protéine avec la surface, le signal de diffusion est capté et mesuré par l'instrument (A). Ce signal localisé est indépendant des autres (B) et peut-être traité séparément (C-D). L'intensité du signal est directement proportionnelle à la masse de la particule. Les distributions en D sont colorées en fonction des taches en C.

### 1.3.6 Spectrométrie de masse MALDI-TOF

La spectrométrie de masse est une technique développée au début du XXI<sup>ème</sup> siècle par J. J. Thomson qui recevra un prix Nobel de physique en 1906. Un deuxième prix Nobel de chimie sera attribué à Francis W. Aston en 1922 pour les améliorations qu'il apportera à la technique.



Les deux premières étapes de la spectrométrie de masse sont d'abord d'ioniser l'échantillon puis de séparer les composés selon le rapport de leur masse sur leur charge ( $m/z$ ) via un analyseur. L'une des méthodes d'ionisation est la désorption-ionisation laser assistée par matrice (MALDI, Matrix assisted laser desorption ionisation). Elle est souvent couplée à un analyseur à temps de vol (TOF, time-of-flight).

L'échantillon est mélangé à une matrice qui va améliorer ses propriétés d'ionisation. Le mélange est ensuite déposé sur un support métallique où un laser UV va désorber et ioniser les molécules. Cette méthode génère principalement des ions monochargés. L'analyseur TOF va lier le temps de vol des ions avec leur rapport  $m/z$ . Les ions sont d'abord accélérés puis leur temps de vol est mesuré. Les ions au petit rapport  $m/z$  auront un temps de vol plus court. Dans les appareils disposant d'un tube de vol la séparation est dépendante de l'accélération lors du passage dans l'accélérateur.

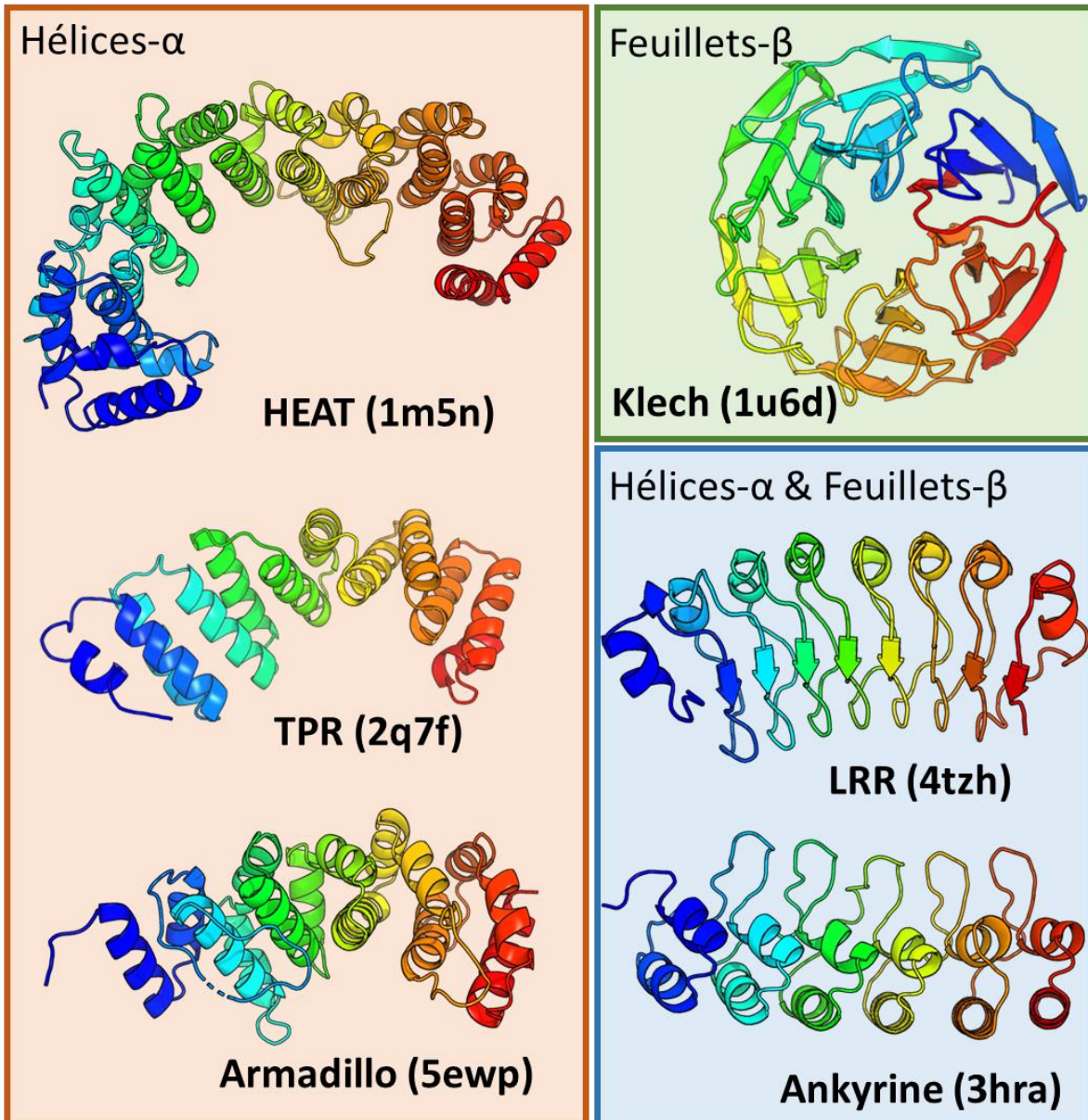
Cette méthode est très peu gourmande en quantités et demande entre 1 à 10 $\mu$ g de protéines.

## 1.4 Alphareps

### 1.4.1 Alphareps origine

Nous avons vu que les anticorps sont les principales macromolécules utilisées en ingénierie protéique en ce qui concerne les interactions spécifiques de fortes affinités. Les immunoglobulines dominent aussi très largement le marché pharmaceutique autant en diagnostic que comme médicament. À ceci s'ajoute le fait qu'elles ont longtemps été le seul outil remplissant une telle fonction, ce qui explique qu'elles aient bénéficié d'autant d'années de recherches et de développement. Cela nous amène aujourd'hui à un catalogue extrêmement diversifié tant du point de vue des applications que des variantes topologiques (figure 1.2.5-1). Cependant, leur utilisation est associée à des contraintes majeures notamment liées aux conditions d'expression contraignantes requises pour leur production.

C'est pour pallier ces limitations que diverses protéines artificielles ont été mises au point. La stratégie commune fut donc de fournir une alternative stable, s'exprimant facilement et ayant la propriété si convoitée des anticorps à pouvoir se fixer de façon spécifique et avec la meilleure affinité à un partenaire identifié. Très rapidement les recherches ont convergé vers les protéines naturelles à répétitions. Ces protéines aux rôles divers sont spécialisées dans les interactions entre protéines dans la cellule. Il est possible de classer ces protéines selon leurs caractéristiques structurales (figure 1.4.1-1). Certaines sont exclusivement composées d'hélices alpha (Armadillo, HEAT et TPR-like), d'autres de feuillets (Klech) ou bien une alternance de feuillets et d'hélices (LRR et Ankyrine) (Andrade et al. 2001a; Grove et al. 2008).



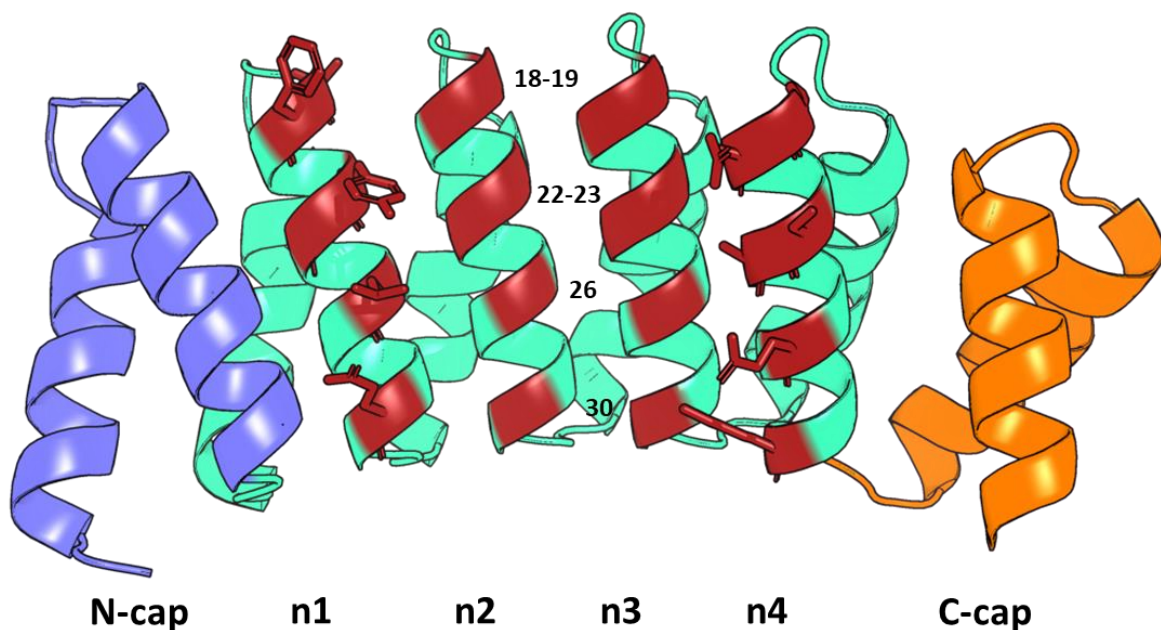
**Figure 1.4.1-1.** Aperçu de différentes familles de protéines à répétitions classées selon leur composition en hélices- $\alpha$  et feuillets- $\beta$ . Le code PDB associé pour chaque structure est indiqué entre parenthèses.

### 1.4.2 L'architecture des alphareps

Les alphareps sont publiées pour la première fois en 2010 (Urvoas et al. 2010). Ce sont les premières protéines artificielles conçues à partir du motif de répétition HEAT. Avant elles, les Darpins sont les pionnières du domaine avec l'utilisation du motif Ankyrine (Mosavi et al. 2002; Binz et al. 2003). HEAT est l'acronyme des quatre familles originelles dans lesquelles il a été identifié : Huntingtin, elongation factor 3 (EF3) 1, protein phosphatase 2A (PP2A) 2, et PI3-kinase TOR1 (Andrade and Bork 1995; Andrade et al. 2001b). Le motif HEAT est formé par deux hélices alpha antiparallèles reliées par une boucle. Cette unité de base est répétée plusieurs fois successivement et adopte une conformation solénoïde typique (figure 1.4.2-1).

Chez les alphareps nous retrouvons cette même structure tertiaire. À partir d'alignement de séquences provenant d'organismes thermophiles il a été possible d'identifier une séquence consensus de 31 acides aminés pour constituer la brique de base de l'unité répétée. De même, deux séquences constantes appelées N-cap et C-cap ont été conçues pour débiter et clôturer l'assemblage de l'alpharep. Ce savant assemblage permet aujourd'hui l'obtention d'alpharep particulièrement stables en plus d'une expression optimale en bactérie.

La variabilité des séquences à certaines positions est une propriété intrinsèque à cette famille de protéines. Une partie sont associées au maintien de la cohésion inter-chaînes et ne peuvent donc pas être mutées sans une répercussion sur la stabilité de l'assemblage. En revanche certaines positions variables (18, 19, 22, 23, 26 et 30) localisées sur la face externe de la deuxième hélice de l'unité (figure 1.4.2-1) sont les candidats parfaits pour la création d'une région d'hypervariabilité aléatoire, à la manière des région CDRs chez les mABs. C'est donc la face concave de l'alpharep qui est engagée dans l'interaction spécifique.



**Figure 1.4.2-1.** Organisation structurale de base d'une alpharep. En violet le N-cap, en orange le C-cap. Les unités répétées sont colorées en cyan-vert (n1-4). Les acides aminés des régions d'hypervariabilité sont colorés en rouge, et affichés en représentation bâton sur les unités 1 et 4.

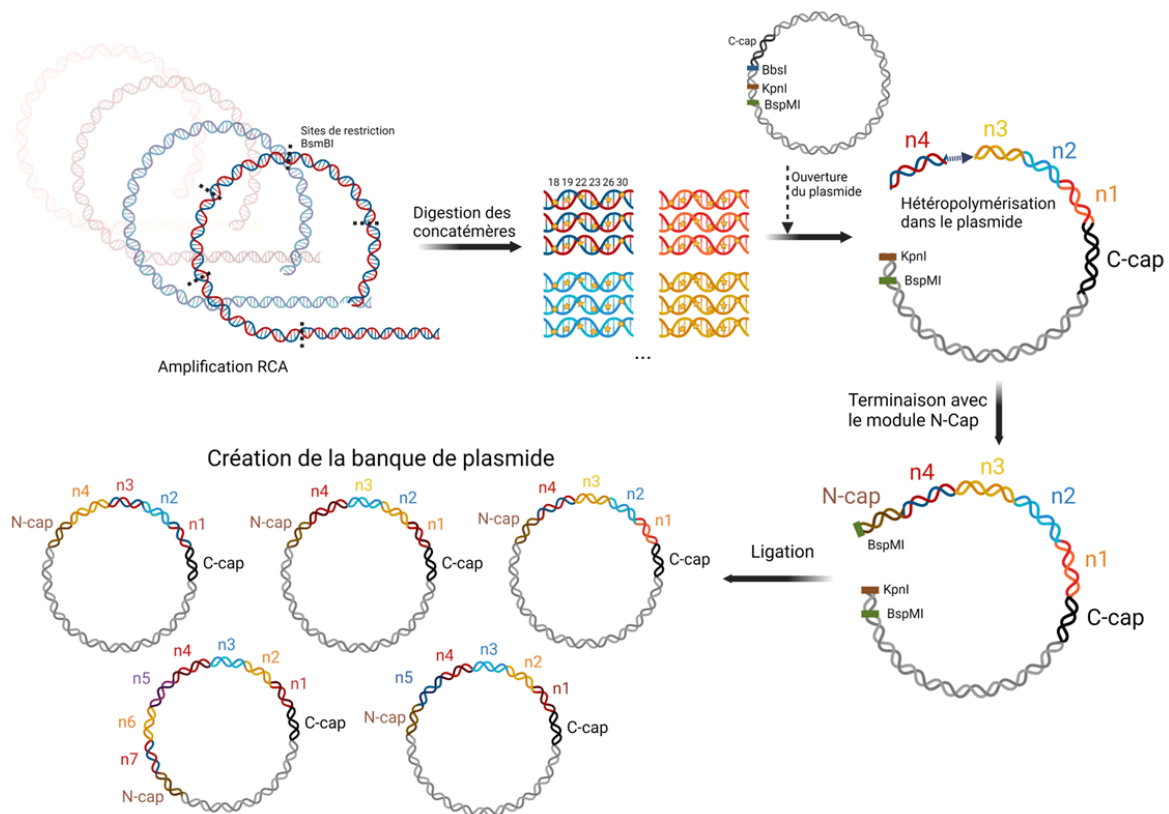
Séquence consensus pour le N-cap : ADPEKVE MYIKNLQDDSDMVRNTAATALGKI

Séquence consensus de l'unité répétée : GDERAVEPLIKALKDEDXXVRXXAAXALGX I

Séquence consensus pour le C-cap : GGERVRAAMEKLAETGTGFARKVAVNYLETH

### 1.4.3 Design de la bibliothèque et création de la diversité

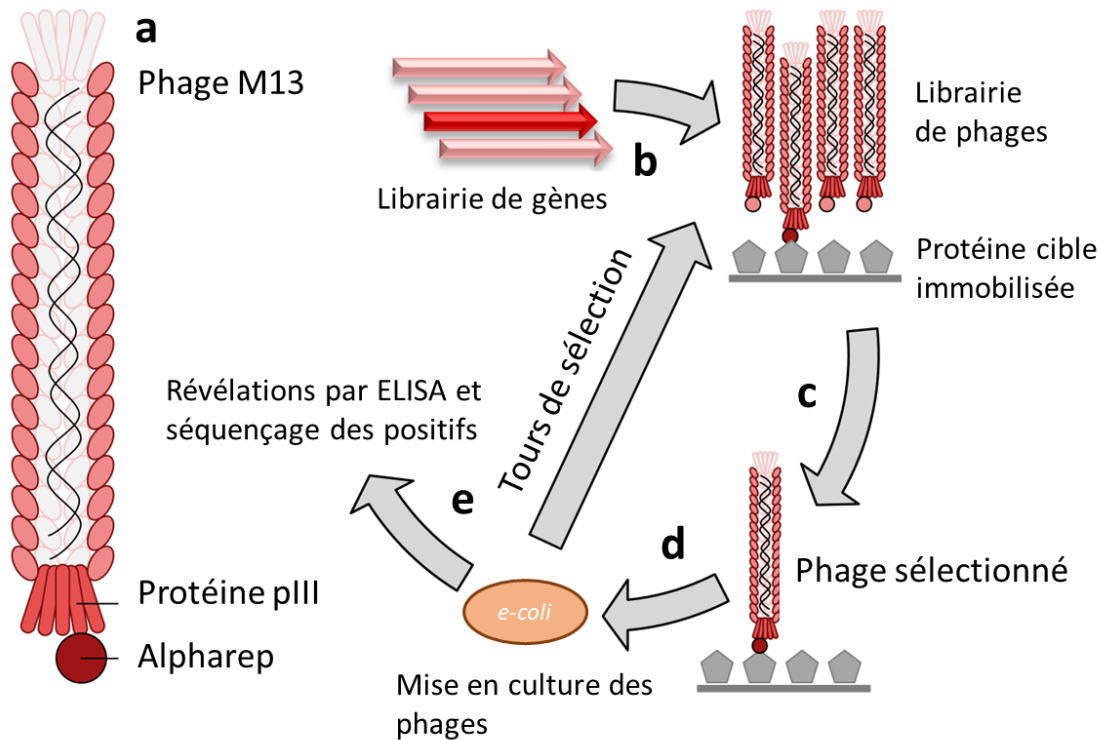
Une sélection d'acides aminés a été faite pour chaque position variable. Certains acides aminés sont rigoureusement exclus tels que les cystéines et les prolines (excepté en position 18) afin d'éviter de perturber le repliement et la stabilité. Ensuite, pour chaque position, les acides aminés candidats sont choisis en fonction de leur fréquence dans les séquences naturelles, avec un enrichissement en tyrosines et tryptophanes. Cette diversité est retranscrite dans une banque ADN (micro-gène) synthétisée puis amplifiée par réplication *rolling circle* (RCA) pour obtenir des homo-oligomères d'ADN circulaires (concatémères). Ces concatémères ont ensuite été digérés pour libérer la multitude de copies de chaque séquence dégénérée (figure 1.4.3-1). Enfin les séquences ont été hétéro-polymérisées aléatoirement et avec une longueur variable de 0 à 9 unités répétées pour créer une banque diversifiée de séquences d'alpharep uniques. Cette stratégie est directement à l'origine de la conception de la première banque apparue en 2010 qui a ensuite été drastiquement optimisée pour encore l'enrichir en diversité (Guellouz et al. 2013). L'actuelle diversité de la librairie 2.1 est estimée à  $1,7 \times 10^9$  alpharep, qui correspond au nombre de grains de sable dans  $0,017 \text{m}^3$  soit 17 litres ou bien le nombre de grains de riz de 34 mille paquets de riz thaï de 1kg (1 grain  $\sim 0,02\text{g}$ ). Heureusement la sélection se fait par phage display, à partir de phage M13 de taille de quelques nanomètres, ce qui permettra d'exprimer l'intégralité de la diversité dans un tube à essai.



**Figure 1.4.3-1.** Image adaptée de Uvoas et al. 2010 (Urvoas et al. 2010). Illustration simplifiée des grandes étapes menant à la diversification de la banque d'alphareps. Les oligonucléotides codant pour les 31 acides aminés et incluant les différentes combinaisons de mutations sont amplifiés par répllication circulaire (RCA). Les concatémères sont ensuite digérés par BsmBI pour libérer les différentes unités. Le plasmide accepteur est ouvert par digestion avec BspMI et KpnI qui permet de lancer une polymérisation orientée coté C-cap et empêcher les hybridations coté KpnI. Les répétitions sont polymérisées dans le vecteur pour former un enchaînement aléatoire d'unités uniques. La polymérisation est terminée par l'ajout d'un excès de N-cap et de la digestion du plasmide et du N-cap avec l'enzyme BspMI, suivie d'une ligation pour fermer le plasmide.

#### 1.4.4 Phage display

La sélection des alphareps se fait par la technique de phage display (Marintcheva, 2018). Dans les grandes lignes, cette technique repose sur les propriétés particulières des phages M13 à exprimer une protéine à leur surface, dans notre cas une alpharep. La protéine d'intérêt est fixée sur une surface puis les phages sont sélectionnés selon l'affinité de l'alpharep exprimée à leur surface pour la protéine cible. Il est ensuite possible de sélectionner des phages uniques suite à la mise en culture de bactéries (*Escherischia coli*) infectées qui sont isolées par étalement sur boîte de Petri. Chaque clone de bactérie est alors infecté par un phage unique. Une révélation par test ELISA permet d'identifier les candidats à plus haute affinité. Il est ensuite possible de séquencer l'alpharep portée par le vecteur du phage (phagemid) (figure 1.4.4-1), alpharep qu'il sera ensuite possible d'exprimer en culture cellulaire *E.coli*.

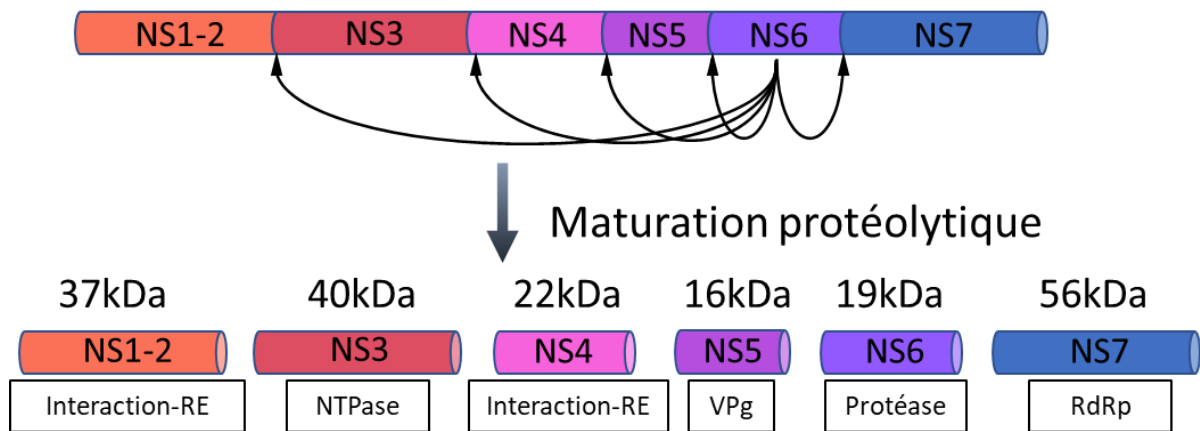


**Figure 1.4.4-1.** Image modifiée de Thomas Shafee 2019 wikimédia. Processus de sélection par phage display. Les phages M13 expriment les alphareps à leur surface (a). La librairie de gènes est clonée dans les phages M13 (b). La protéine est immobilisée sur la surface et les phages sont mis en contact. L'alpharep exprimée à la surface des phages se fixe à la cible selon son affinité permettant d'enrichir en phages liés (c). Le stock de phage est régénéré via une culture dans des bactéries. Le cycle correspond à un tour de sélection et est répété plusieurs fois. Après plusieurs tours, les bactéries sont isolées puis l'affinité d'une sélection aléatoire de clones est estimée par ELISA. Les gènes des clones positifs sur la révélation ELISA sont séquencés et clonés dans un vecteur d'expression E-coli pour lancer la production d'alpharep en E-coli (e).

## 1.5 NS4

J'ai été amené à travailler sur un autre projet que le complexe HTP, portant sur une petite protéine de 22 kDa associée aux membranes, NS4. NS4 est une protéine de Norovirus libérée au cours du processus de maturation de la polyprotéine ORF1 (Habersetzer *et al.*, 2020). L'infection par les norovirus est la cause la plus courante de gastro-entérite virale aiguë dans le monde. Les norovirus sont divisés en sept génogroupes de GI à GVII, eux-mêmes subdivisés en plusieurs génotypes. Les génogroupes I, II et IV infectent principalement les humains, le génotype 4 du génogroupe II (GII.4) étant à l'origine de la majorité des infections actuelles (Tran *et al.*, 2013). Le génome des norovirus, en particulier ces norovirus humains (huNoV) est composé de trois cadres de lecture ouverts (ORF). Les ORF2 et ORF3 codent pour des protéines "structurales" (de la capsid), tandis que l'ORF1 code pour une polyprotéine (environ 190 kDa) qui contient plusieurs domaines qui sont maturés en protéines "non structurales", *i.e.* qui ne sont pas incorporées dans les particules virales mais sont impliquées dans la réplication virale intracellulaire : NS1-2, NS3, NS4, NS5, NS6, NS7 (Figure 1.5-1). La protéine NS4 est la moins

bien caractérisée. Les études ont montré son implication dans le remodelage des membranes du réticulum endoplasmique, supposément en relation avec la protéine NS1-2 (Doerflinger et al., 2017; Hung et al., 2023).



**Figure 1.5-1.** Illustration tirée de Habersetzer et al. 2020 (Habersetzer et al. 2020). La protéine NS4 du norovirus humain provient de la polyprotéine ORF1. La maturation protéolytique assurée par NS6 libère les six produits NS1 à 7. L'encadré sous les protéines indique quelques-unes des fonctions connues pour chaque protéine. RE = Réticulum endoplasmique, NTPase = Nucléoside triphosphatase, VPg = Viral protein genome-linked, RdRp = RNA-dépendent RNA polymerase.

## 2 Structure du complexe sHER2-Pertuzumab-Trastuzumab Fabs

Le projet de thèse est axé sur la méthodologie autour de la préparation de complexes anticorps-antigène et leur analyse en particules isolées (SPA) par microscopie électronique à transmission. Dans ce chapitre nous décrirons nos observations sur le complexe sHER2-Pertuzumab-Trastuzumab depuis la biochimie (pré-microscope), à la préparation de la grille et son observation (in-microscope) et au traitement des données (post-microscope) qui feront office de sous-parties à ce chapitre.



**Pre-microscope :** Une vue biochimique et biophysique de l'échantillon

**In-microscope :** Préparation de grilles et approches optimisée pour un screening de grilles efficient. La quête de la grille ultime.

**Post-microscope :** Le traitement des données





## 2.1 Pre-microscope : Une vue biochimique et biophysique de l'échantillon

### 2.1.1 Matériel et méthodes : purification, analyses biophysiques

#### 2.1.1.1 Préparation de l'ectodomaine de la protéine humaine HER2

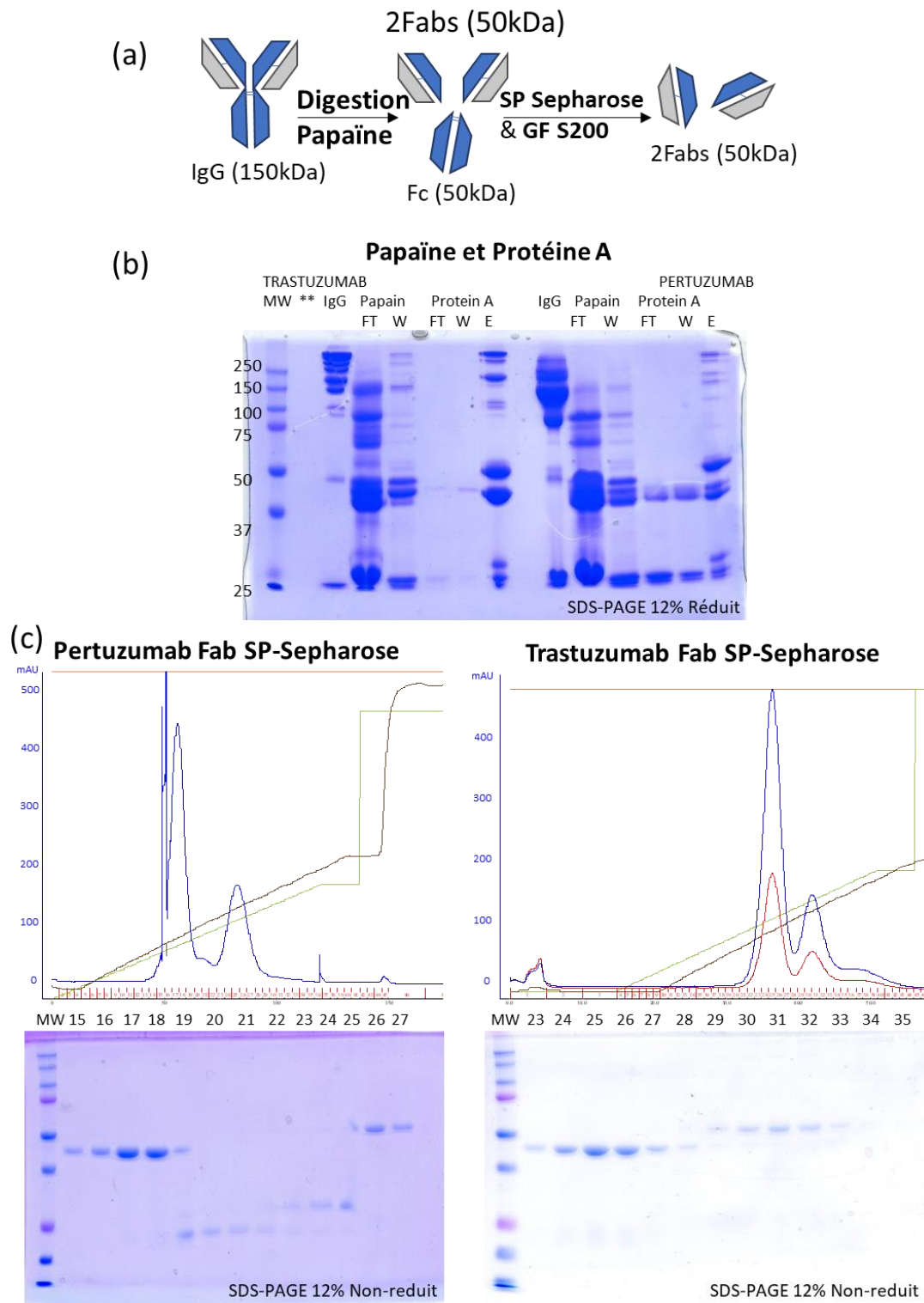
Pour réaliser les travaux de la thèse nous avons utilisé l'ecto-domaine de la protéine humaine HER2 comprenant les résidus Thr23-Thr652 selon la numérotation de la protéine entière avec un marqueur poly-histidine en C-terminal ci-après dénommé sHER2. Dans la suite de la thèse la protéine sera numérotée de 1 à 629 et dénommée sHER2. Nous avons commandé la protéine sous forme lyophilisée chez Acrobiosystems inc (Cat. No. #HE2-H5212). Pour sa purification elle est reprise dans un tampon HEPES pH7,5, 150 mM NaCl et injectée sur une colonne superdex 200 increase 10/300 GL à 7°C. Les fractions qui contiennent sHER2 sont concentrées à une DO de 3,2 et réparties en aliquotes de 15 µl avant une congélation rapide dans l'azote liquide et un stockage à -80°C.

#### 2.1.1.2 Purification de la région "fragment antigen-binding (Fab)" à partir des anticorps-monoclonaux (mABs) du pertuzumab et du trastuzumab

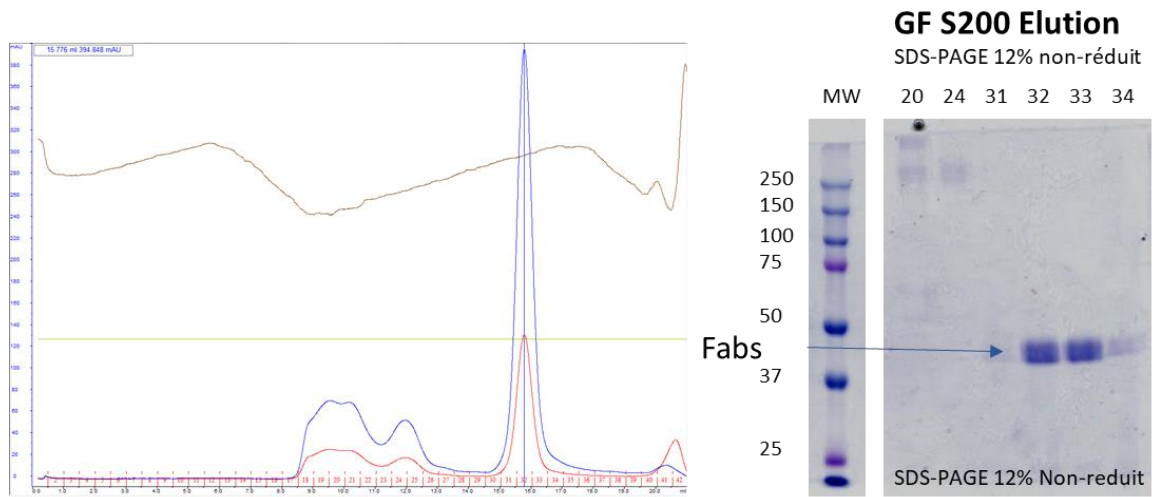
Les anticorps monoclonaux (mABs) ont été achetés chez Clinisciences (Nanterre, France, Cat. No. HY-P9912 et HY-P9907). Selon les informations du fournisseur, le pertuzumab et le trastuzumab sont des anticorps monoclonaux humanisés produits par la technique de l'ADN recombinant dans une culture de cellules de mammifère CHO (ovaire de hamster chinois) contenant l'antibiotique gentamicine. Le pertuzumab et le trastuzumab ont un poids moléculaire approximatif de 148 kDa.

Ils sont solubilisés dans un tampon phosphate de sodium à 20 mM et pH 7,5 à 3 mg/ml (DO~4,2) puis stockés à -20°C en aliquotes de petits volumes congelés rapidement dans l'azote liquide. La digestion des mABs pour produire les Fabs est faite avec de la papaine immobilisée commerciale (#20341 Thermo-Fisher). 250 µl de papaine immobilisée à 50 % sont déposés sur une colonne centrifugeable de 1ml (#69725 Thermo-Fisher) puis lavés avec de l'eau miliQ avec une centrifugeuse de pailasse. Ensuite la papaine est équilibrée et activée avec le tampon de digestion à 20 mM de cystéine, 10 mM d'EDTA, 20 mM de phosphate de sodium à pH7,5. 20 mM de cystéine sont ajoutés à l'anticorps. La papaine activée et les anticorps sont mélangés dans un tube Eppendorf de 2ml. La digestion s'effectue à 37°C pendant 4-5 heures sous agitation constante avec un intelli-mixer afin d'éviter la sédimentation de la résine. Ensuite l'élution est faite par centrifugation à 1000g. La séparation du Fab et du Fc se faisant mal par passage sur protéine A, ils sont donc séparés par une chromatographie échangeuse d'ions SP-sepharose de 3 ml (#17-0729-10 GE-Healthcare) (figure 1.1.2-1a et 1b). Avant l'injection, les mélanges sont dialysés sur la nuit à 7°C dans un tampon à 50 mM d'acétate à pH 4 puis ils sont injectés sur la SP-sepharose. Le chargement et le lavage sont faits dans le même tampon que la dialyse, puis l'élution dans un gradient de 40 % de tampon B de 50 mM acétate pH 4 plus 1M NaCl en 120min à 0,3ml/min. Le premier pic d'élution contient le Fab et le second le Fc et les IgG non digérés (figure 1.1.2-1c). Les fractions contenant les Fabs sont rapidement

neutralisées en ajoutant un faible volume d'HEPES 1M pH7. Après vérification par gel SDS-PAGE à 14 % d'acrylamide en condition non réductrice les fractions contenant les Fabs sont dialysées sur la nuit à 7°C dans un tampon HEPES pH7,5 et 150 mM NaCl ou bien par changement rapide de tampon par centrifugation dans un centricon de 30kDa. Les fractions sont mélangées puis concentrées à 1ml avant d'être injectées dans une chromatographie d'exclusion Superdex 200 10/300 GL (SEC Citiva). Les fractions sont déposées sur des gels SDS-PAGE 14 % d'acrylamide l'un en condition réductrice où les bandes des Fabs sortent à ~25kDa et l'un en condition non réductrice (sans beta-mercaptoéthanol) où les bandes sortent légèrement en dessous du marqueur des 50kDa (figure 1.1.2-2). Les fractions sont rassemblées et concentrées à une DO de 3,2 avant d'être divisées en aliquotes de 15 µl qui sont stockés à -80°C après congélation rapide dans l'azote liquide.



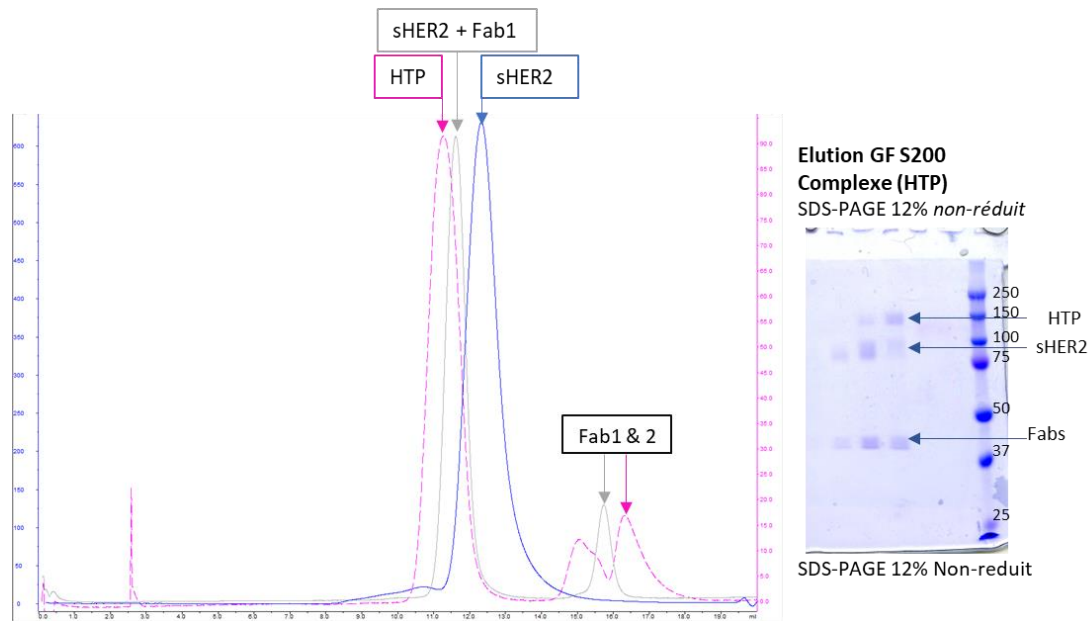
**Figure 1.1.2-1.** (a) schéma du processus de préparation des Fabs. (b) Gel SDS-PAGE des résultats du traitement par la papaïne et le passage sur protéine A : MW = Marqueur de poids moléculaire (Molecular Weight), FT= Volume non retenu (Flow through); W= Lavage (wash); E= Elution. (c) Les profils d'éluion du passage sur SP-sépharose et les gels SDS-PAGE correspondant en dessous pour chacun des anticorps. Les courbes bleues montrent l'absorbance à 280 nm et la courbe rouge à 254 nm.



**Figure 1.1.2-2.** Profil d'élution des Fabs par chromatographie d'exclusion S200. À gauche le profil d'élution de la S200 à droite le gel SDS-PAGE 12 % d'acrylamide correspondant.

### 2.1.1.3 Création du complexe

Le complexe a été formé en incubant les deux Fabs issues du pertuzumab et du trastuzumab avec sHER2. Le procédé est effectué dans la glace durant quelques minutes (durée variable), avec un ratio de 1:1,2 sHER2/Fabs. Le complexe est ensuite purifié sur une chromatographie d'exclusion stérique dans une colonne de Superdex 200 increase 10/300 GL dans 10 mM HEPES pH 7,5 et 150 mM NaCl. Un déplacement du volume d'élution entre sHER2 seul et sHER2 après incubation avec les Fabs est visible, correspondant à une augmentation de la masse et donc à la formation du complexe (figure 1.1.3.1). Le complexe est suffisamment stable pour être visible en SDS-PAGE 10 % acrylamide et 0,1 % SDS en condition non réductrice.



**Figure 1.1.3-1.** Présentation d'un chromatogramme typique obtenu lors de la purification du complexe HTP. A Gauche le profil d'élution à 280 nm et à droite le gel SDS-PAGE correspondant.

#### 2.1.1.4 Spectre UV

Environ 2-2,5 µl de l'échantillon est déposé sur la cellule du nano-drop spectrophotomètre (nanodrop one thermofisher scientific) et un spectre UV est enregistré de 230 nm à 350 nm. Les données sont exportées en format excel pour être traitées. L'indice d'agrégation est calculé par la formule :

$$A_i = \frac{100 * A_{340}}{A_{280} - A_{340}} < 2$$

Il est considéré comme acceptable si inférieur à 2.

#### 2.1.1.5 DLS et nano-DSF

L'échantillon est utilisé à une densité optique entre 0,3 et 1 puis centrifugé à 11 000g pendant 10 minutes. Environ 10 µl sont utilisés pour remplir le capillaire. Pour la DLS les mesures sont faites entre 15°C et 25°C aucune différence n'est visible à l'une ou l'autre température. Les mesures de nano-DSF sont faites sur une gamme de températures allant de 15°C à 95°C avec un pas de 0,5°C. Les données sont traitées avec le logiciel nanotemper ou excel.

#### 2.1.1.6 Photométrie de masse par Refeyn

La masse et l'abondance relative sont mesurées avec un TwoMP mass photometer (Refeyn Inc., Oxford, UK). Un échantillon congelé depuis 1 mois a été décongelé avant d'être déposé sur la cellule de mesure du Refeyn. L'échantillon est chargé sur le support en silicone (CultureWell reusable gasket, 3mm de diamètre et 1 mm de profondeur, Grace Bio-Labs) puis placé sur le support en verre propre. Chaque puits est rempli avec 10 µl de tampon (10 mM MES pH6,5, 150 mM NaCl) pour faire le blanc. Ensuite, la protéine est ajoutée à la goutte de tampon puis la suspension est homogénéisée pour une concentration finale de 10 nM. Le poids moléculaire de l'échantillon est retrouvé à partir d'une gamme étalon effectuée le jour même ou auparavant. La gamme étalon était composé de la BSA (66,5 kDa + 133 kDa), une IgG (~150 kDa) et une Thyroglobuline (660 kDa).

#### 2.1.1.7 Chromatographie d'exclusion de taille couplée à la mesure de diffusion statique de la lumière (SEC-SLS).

L'échantillon fraîchement purifié est analysé par SEC-SLS à l'institut pasteur. Pour cela 50 µl d'échantillon est injecté et analysé avec un OMNISEC (Malvern Panalytical) avec une Superdex 200 10/300 GL équilibrée dans 10 mM HEPES pH 7,5 et 150 mM NaCl couplée à un détecteur de diffusion statique de la lumière (SLS), un détecteur UV. Le détecteur d'indice de réfraction n'a pas pu être utilisé.

## 2.1.2 Résultats : une vue exhaustive du comportement du complexe par les méthodes biophysique

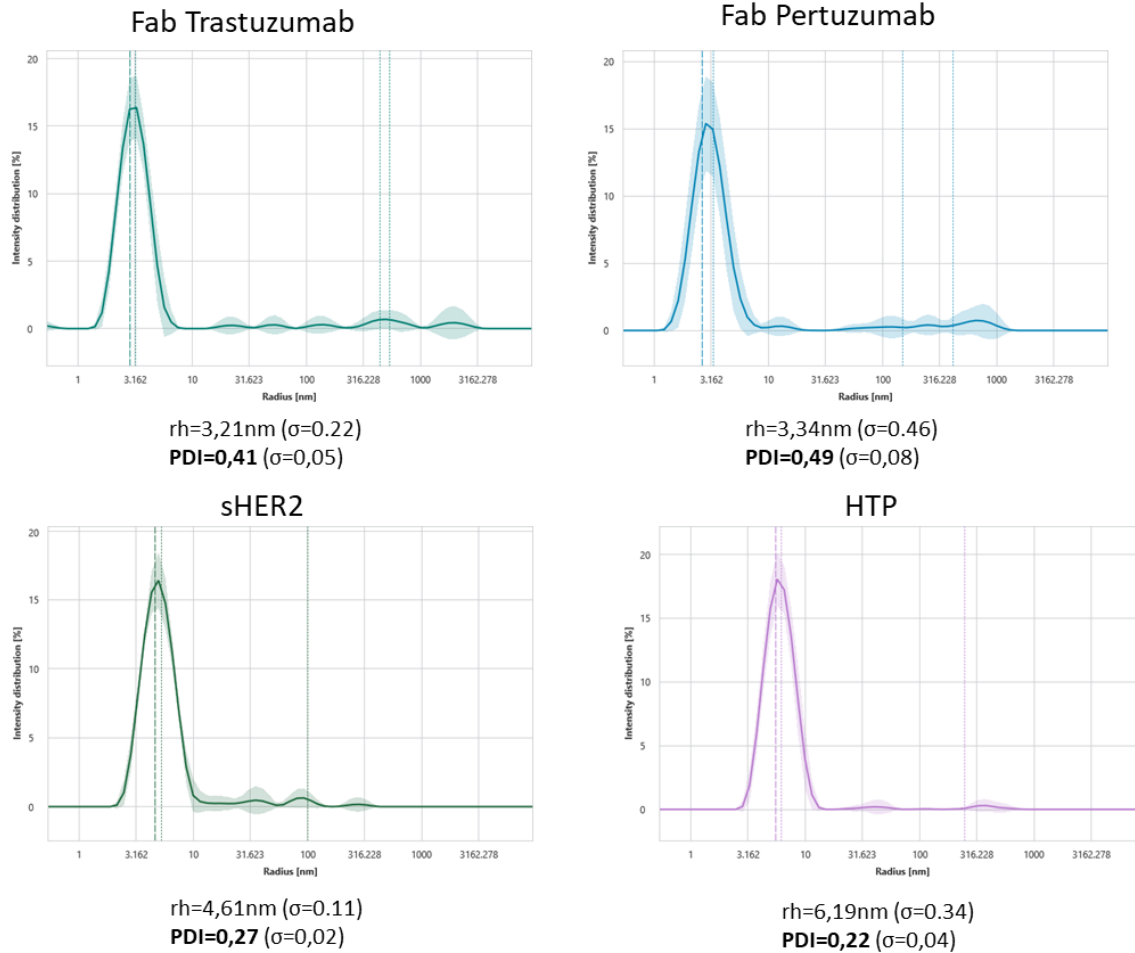
Le premier point est la vérification de la bonne formation du complexe. La technique la plus utilisée est sans doute la chromatographie d'exclusion stérique sur FPLC. Les fragments d'anticorps sont des partenaires d'environ 50kDa connus pour leur grande stabilité. Ils présentent des profils d'élution particulièrement nets ce qui permet de facilement visualiser lorsque le complexe se forme. En plus de permettre un contrôle de la formation du complexe anticorps-antigène, la FPLC permet d'homogénéiser la préparation en isolant le complexe des partenaires libres et en retirant les éventuels agrégats. Ainsi les échantillons ayant un profil correct peuvent directement être utilisés pour préparer des grilles de cryo-EM.

Nous pouvons distinguer deux ensembles de mesures : un premier ensemble qui est effectué en routine avant la préparation des grilles il comprend la DLS, la nano-DSF et le spectre UV, et un second qui inclut des mesures ponctuelles sur des échantillons ayant été préparés uniquement pour enrichir la caractérisation biophysique du complexe HTP.

### 2.1.2.1 Contrôle de routine par spectre UV, DLS & nano-DSF

Les mesures du spectre UV, de DLS et de nano-DSF sont réalisées en routine chaque jour où l'échantillon est passé sur FPLC puis préparé pour la cryo-EM. Ci-après dans ce paragraphe seront présentées les mesures considérées par la littérature comme standards pour le complexe HTP, à savoir dans un tampon HEPES à pH 7,5 et avec 150 mM de NaCl.

En plus de l'absolument nécessaire mesure de la concentration de l'échantillon (280 nm) le spectre UV renseigne sur la présence de contaminations par le ratio d'absorbance 280/260 nm ainsi que sur l'indice d'agrégation (cf matériel et méthode). Ensuite, la DLS renseigne non seulement sur la stabilité et l'homogénéité du complexe, mais également sur sa formation. La variation de taille induite par la liaison de 2 Fabs ( $\Delta 100$  kDa) sur sHER2 est suffisante pour induire un changement significatif du coefficient de diffusion mesuré par DLS et donc du rayon hydrodynamique apparent de 3,2 nm ( $\pm 0,22$ ), 3,34 ( $\pm 0,46$ ) et 4,61 nm ( $\pm 0,11$ ) pour le Fab du trastuzumab, le Fab du pertuzumab et sHER2 respectivement, à 6,19 ( $\pm 0,34$ ) pour le complexe ternaire HTP (Figure 1.2-1).

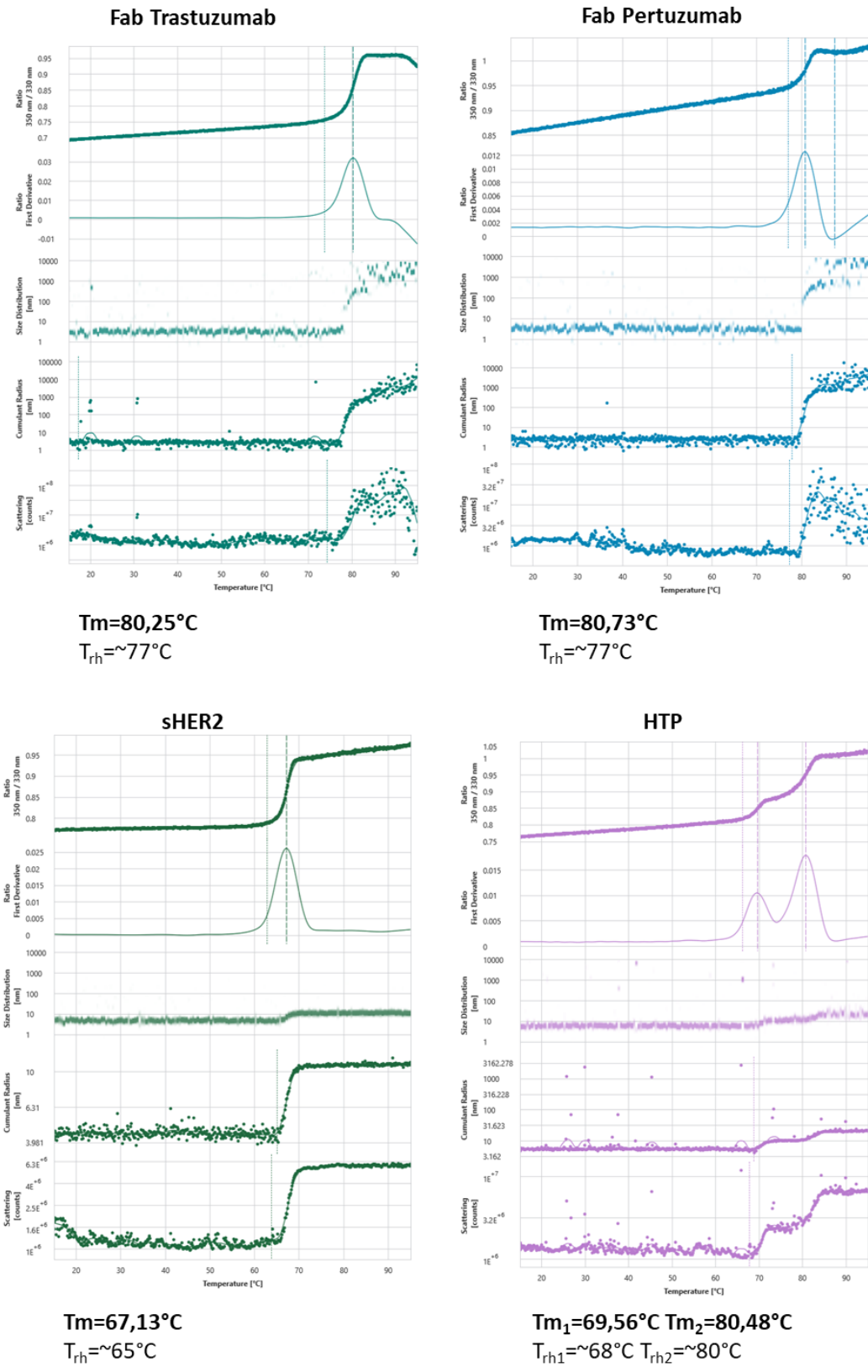


**Figure 1.2-1.** Mesure DLS de tous les partenaires du complexe séparément et du complexe formé. *rh*= rayon hydrodynamique, *PDI*= indice de polydispersité,  $\sigma$  = écart type. Les données DLS sont celles du signal cumulé.

Ensuite, les mesures systématiques de nano-DSF fournissent des informations directes sur la stabilité du complexe en solution par le biais de la température de fusion. Celle-ci peut être déduite des variations de la fluorescence ( $T_m$ ) ou de la diffusion de la lumière ( $T_{rh}$ ) à mesure que la température augmente. Au cours du gradient les données de mesures DLS des Fabs seuls augmentent brusquement à partir des 80°C pour atteindre des valeurs de comptes de diffusion de  $1E^8$  counts et un rayon hydrodynamique (*rh*) proche des 1000 nm très fluctuantes. Pour sHER2, les changements apparaissent dès 65-67°C à partir d'où le compteur d'évènement de diffusion monte à  $6E^6$  counts et le rayon hydrodynamique cumulé indique un signal stable d'environ 10 nm en fin de gradient de température. Ces observations indiquent que les Fabs s'agrègent lors de la dénaturation mais pas sHER2. Concernant les mesures sur le complexe HTP, deux évènements de changement de *rh* ont lieu à 70°C où le *rh* atteint 10 nm avec un compteur d'évènement qui augmente et se stabilise à  $3E^6$  counts puis à 80°C le *rh* évolue à nouveau pour monter à 30-50 nm avec un compteur d'évènement de  $8-9E^6$ . La taille apparente et le compteur d'évènement étant plus élevé pour les Fabs nous pouvons nous en servir pour attribuer les évènements à sHER2 ou aux Fabs.



Les Fabs mesurés seul ont un point de fusion ( $T_m$ ) proche des 80°C, et sHER2 seul a un point de fusion de 67°C. Une fois les trois partenaires mélangés et le complexe formé, le  $T_m$  de sHER2 passe à 69°C, ce qui est attribuable à une stabilisation de sHER2 par les deux Fabs (figure 1.2-2). Le deuxième  $T_m$  attribuable au Fab reste à 80°C. De même pour le gradient de température du complexe HTP les mesures de distribution des tailles démarrent autour des 5-6 nm et terminent proches des 15-16 nm avec de faibles fluctuations, ce qui est bien inférieur à l'observation sur les partenaires isolés. Cette observation est faite avant le passage sur FPLC.



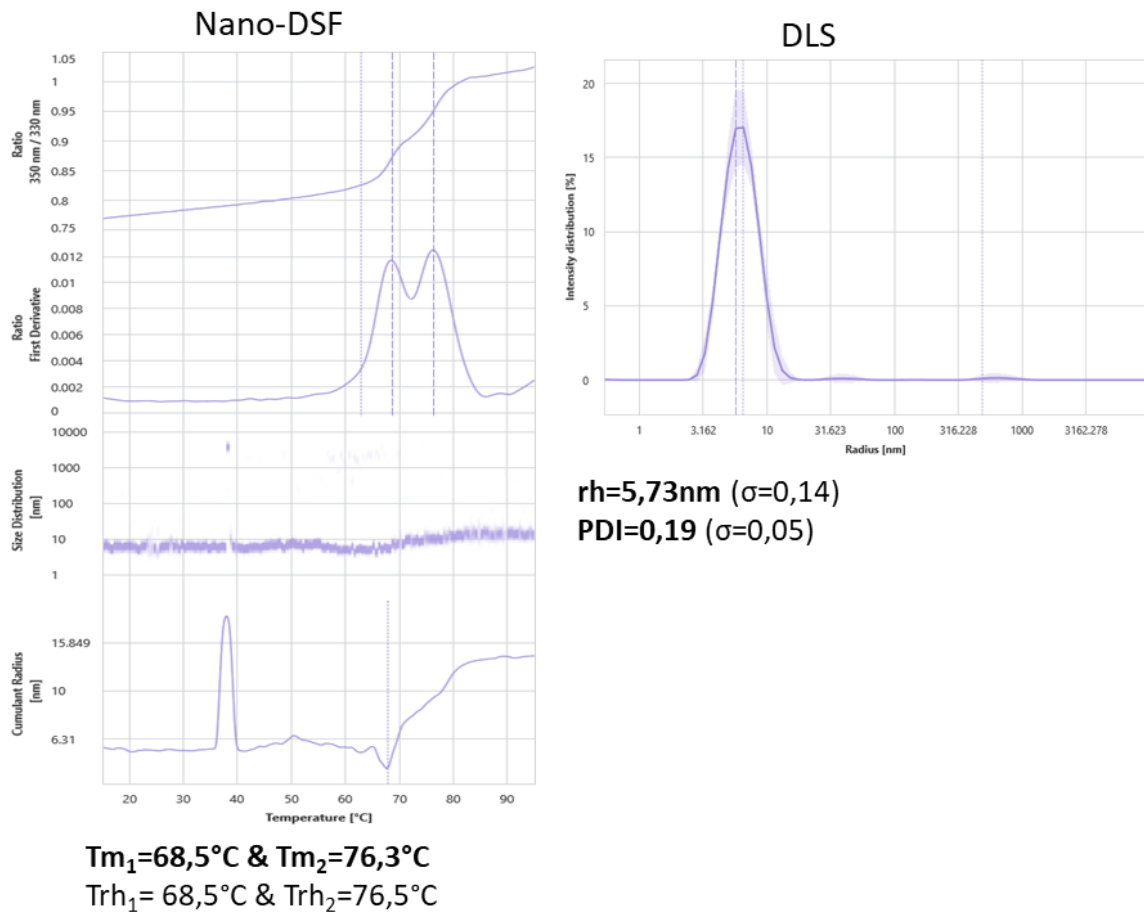
**Figure 1.2-2.** Mesure nano-DSF de chacun des partenaires du complexe individuellement et du complexe HTP dans le tampon standard 10 mM HEPES pH 7,5 et 150 mM NaCl.  $T_m$  = température de fusion,  $T_{rh}$  = température associée aux changements des valeurs de diffusion de lumière.

### *2.1.2.2 Test de stabilité du complexe HTP par DLS et nano-DSF*

Des tests de stabilité face à une variation de pH et de concentration de sel ont été effectués sur le complexe HTP par DLS et nano-DSF. Ces tests ont chronologiquement été réalisés plus tard à la suite de difficultés rencontrées lors de la préparation des grilles de cryo-EM. Cependant afin de simplifier la lecture nous allons les détailler dans ce chapitre.

Les mesures de nano-DSF et DLS montrent des petites variations de la stabilité du complexe à différents pH (figure 1.2-3). À pH 9,2 le  $Tm_1$  est de 68.5°C et le  $Tm_2$  est de 76.3°C. Les mesures DLS simultanées montrent que le rayon hydrodynamique évolue légèrement de 6 nm à 14 nm au cours du gradient de température en un seul évènement et avec un signal homogène et stable. De même la DLS mesurée à 15°C indique un rayon hydrodynamique cumulant proche des 5,73 nm ( $\pm 0,14$ ) avec un indice de polydispersité autour des 0,19 ( $\pm 0,05$ ). Enfin le nombre d'évènement de diffusion commence à  $1,5E^6$  au début du gradient et termine aux alentours de  $5,2E^6$ . L'évolution de la distribution des tailles au cours du gradient de température ainsi que l'indice de polydispersité sont très similaires à pH 9,2 et à pH 7,5. Les températures de fusion ( $Tm$ ) sont légèrement moins bonnes à pH 9 que à pH 7,5 tout comme le rayon hydrodynamique mesuré par DLS qui est légèrement inférieur à celui obtenue à pH 7,5. Ces résultats ont été reproduit avec une autre préparation indépendante. Dans l'ensemble, les valeurs sont relativement proches de celles obtenues à pH 7,5.

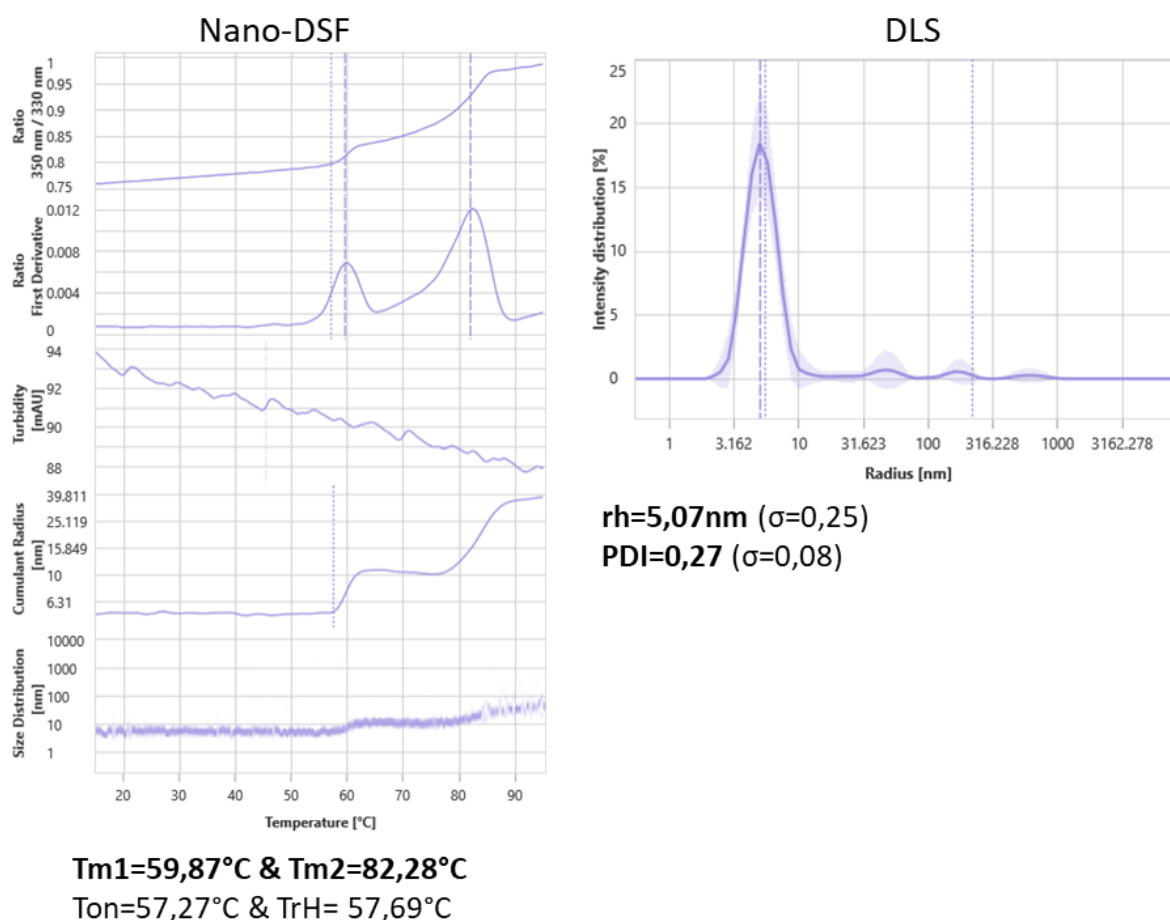
## Glycine pH 9,2 150mM NaCl



**Figure 1.2-3.** Mesure nano-DSF et DLS dans le tampon glycine pH 9,2. Tm= température de fusion, rh= rayon hydrodynamique, PDI = indice de polydispersité. Les données DLS sont celles du signal cumulé.

Lorsque les mesures sont faites à pH 4,6 (figure 1.2-4), on observe une chute du Tm<sub>1</sub> à 59,9°C tandis que le Tm<sub>2</sub> monte à 82,3°C. Quant aux mesures DLS en continu au cours du gradient de température, le saut de rh se produit en deux étapes pour atteindre en fin de gradient des valeurs aux alentours des 100 nm. Un pH très acide semble donc avoir un impact négatif sur sHER2 mais positif sur les Fabs. Enfin les mesures DLS à 15°C donnent des valeurs très variables d'une mesure à l'autre. Le rayon hydrodynamique cumulé oscille entre des 5,13 nm (±0,07) et 7,46 nm (±0,38). Le PDI est entre 0,19 (±0,04) et 0,10 (±0,07). Cette condition semble dans l'ensemble moins favorable à la formation et la stabilité du complexe.

## Acetate pH4,6 150mM NaCl

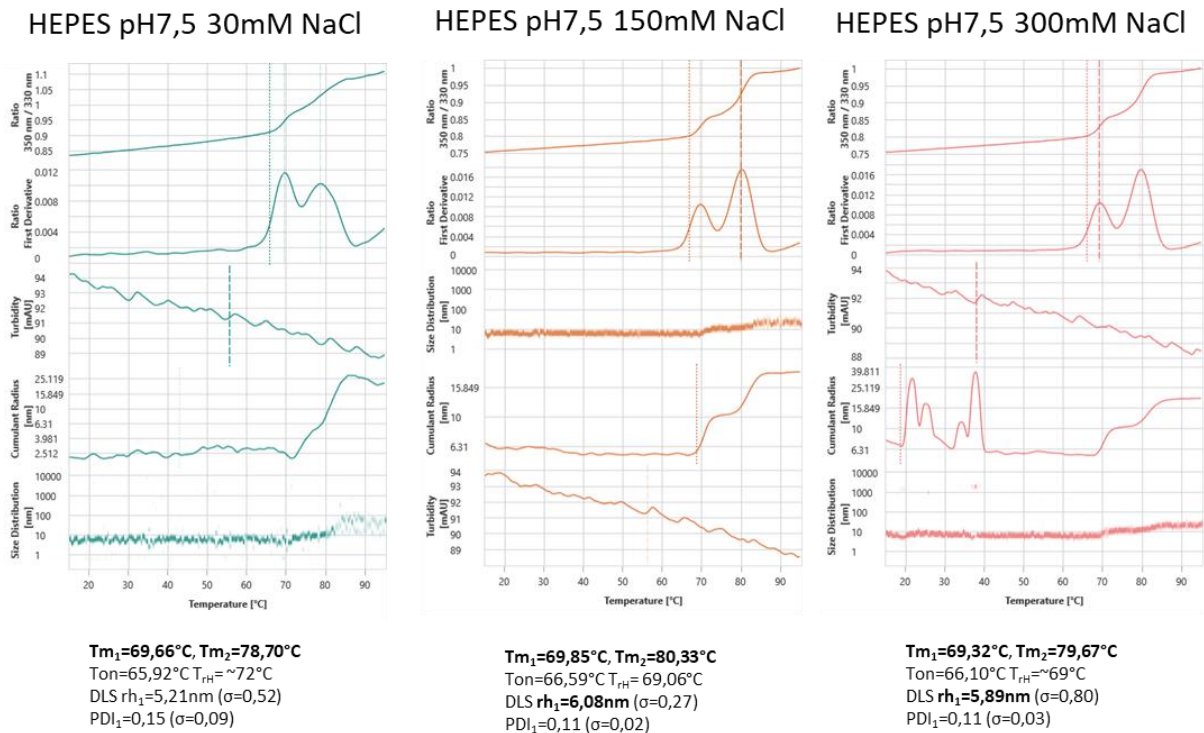


**Figure 1.2-4.** Mesure nano-DSF et DLS dans le tampon acétate pH 4,6. Tm= température de fusion, rh= rayon hydrodynamique, PDI = indice de polydispersité. Les données DLS sont celles du signal cumulé.

Deux autres analyses ont été faites dans un tampon acétate à pH 3,7 et bis-tris pH 6. En résumé, à pH 3,7 la nano-DSF donne des températures de fusion  $Tm_1 \sim 48^\circ\text{C}$  et  $Tm_2 \sim 75^\circ\text{C}$  drastiquement inférieures à celles obtenues dans les autres conditions. De même la DLS montre une chute du rayon hydrodynamique cumulé à  $\sim 4,8 \text{ nm}$  ( $\pm 0,07$ ) avec un PDI cumulé à  $0.19$  ( $\pm 0,04$ ) ce qui est peut-être provoqué par une perturbation de l'assemblage du complexe. Cette observation est la même à différentes concentrations de sel (30 mM NaCl et 150 mM NaCl). Concernant les mesures dans un tampon bis-tris à pH 6, nous sommes proche du point isoélectrique de la protéine ( $pI_{\text{SHER2}}=6,6$ ). Les températures de fusion sont identiques à 150 mM et 30 mM de NaCl et valent  $Tm_1 \sim 64,7^\circ\text{C}$  et  $Tm_2 \sim 82^\circ\text{C}$ . Elles sont bien inférieures à celles obtenues à pH 7,5. Les mesures DLS à  $15^\circ\text{C}$  indiquent un rayon hydrodynamique cumulé de  $4,99 \text{ nm}$  ( $\pm 0,25$ ) et un PDI de  $0.27$  ( $\pm 0,08$ ) à 150 mM et très similaires à 30 mM. Les valeurs sont variables d'une mesure à l'autre.

Une dernière étude de la stabilité du complexe par DLS et nano-DSF à pH 7,5 face à différentes concentrations de sel a été effectuée (figure 1.2-5). En résumé à pH 7,5, l'augmentation du

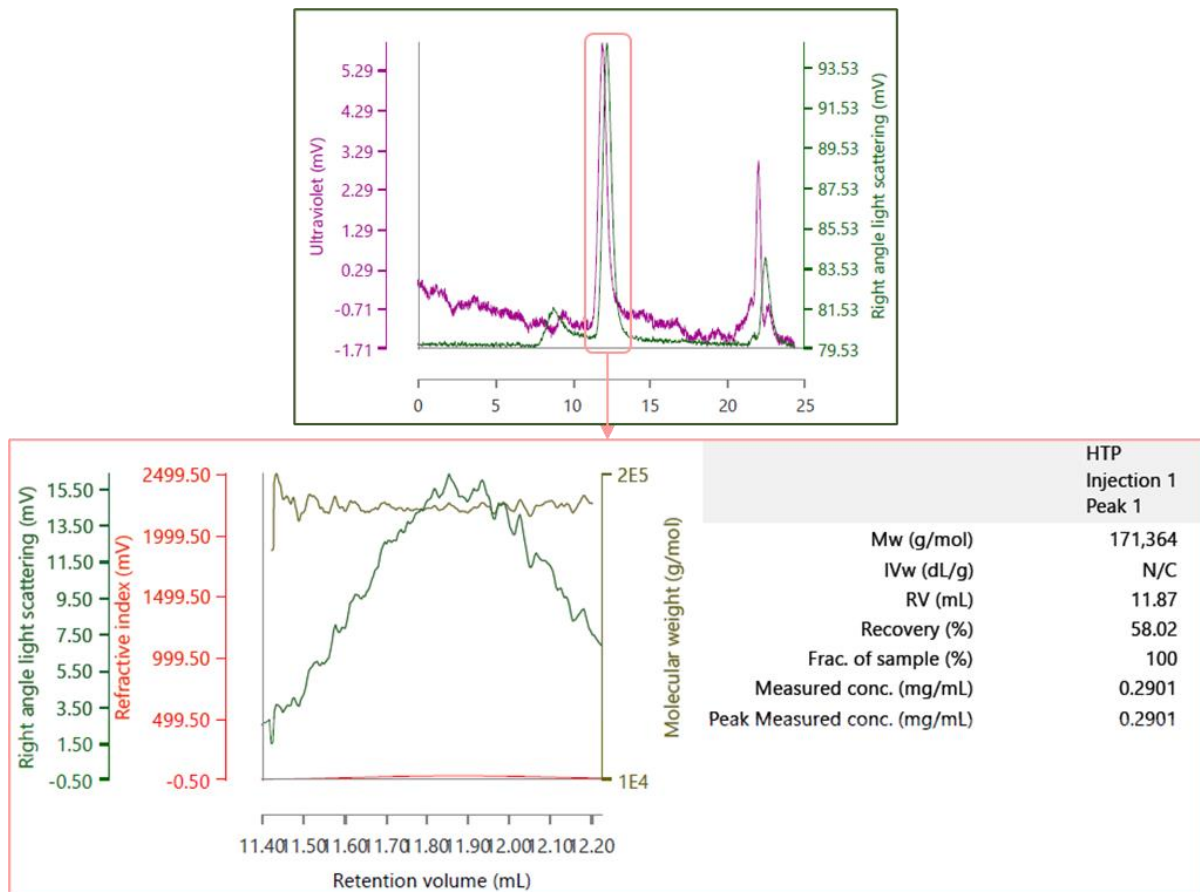
NaCl à 300 mM a peu d'incidence sur les valeurs de  $T_m$  et sur les mesures DLS au cours du gradient de température ou ponctuelle à 15°C. En revanche la diminution de la concentration en NaCl à 30 mM a elle modifié les valeurs de distribution en taille des objets en fin de gradient qui semblent plus gros (~100 nm) et plus fluctuants. Il semble donc que des concentrations de sel dans cette gamme n'impactent que peu la stabilité et l'intégrité du complexe.



**Figure 1.2-5.** Résultat des mesures nano-DSF et DLS dans le tampon à pH 7,5 et à différentes concentrations de sels.  $T_m$  = température de fusion,  $rh_1$  = rayon hydrodynamique du pic 1,  $PDI_1$  = indice de polydispersité du pic 1,  $\sigma$  = écart type.

### 2.1.2.3 Caractérisation par SEC-SLS du complexe ternaire

Les mesures de SEC-SLS dans le pH standard HEPES pH 7.5 et 150 mM NaCl donne une mesure précise de la taille de l'objet et révèle sans ambiguïté une masse apparente de ~171kDa à une concentration en sortie de SEC de 0,3 mg/ml (1.8  $\mu$ M) (figure 1.2-6) et donc très proche du poids moléculaire du complexe HTP (Figure 2.2-3). L'indice de réfraction n'a pas pu être mesuré le capteur étant hors-service.



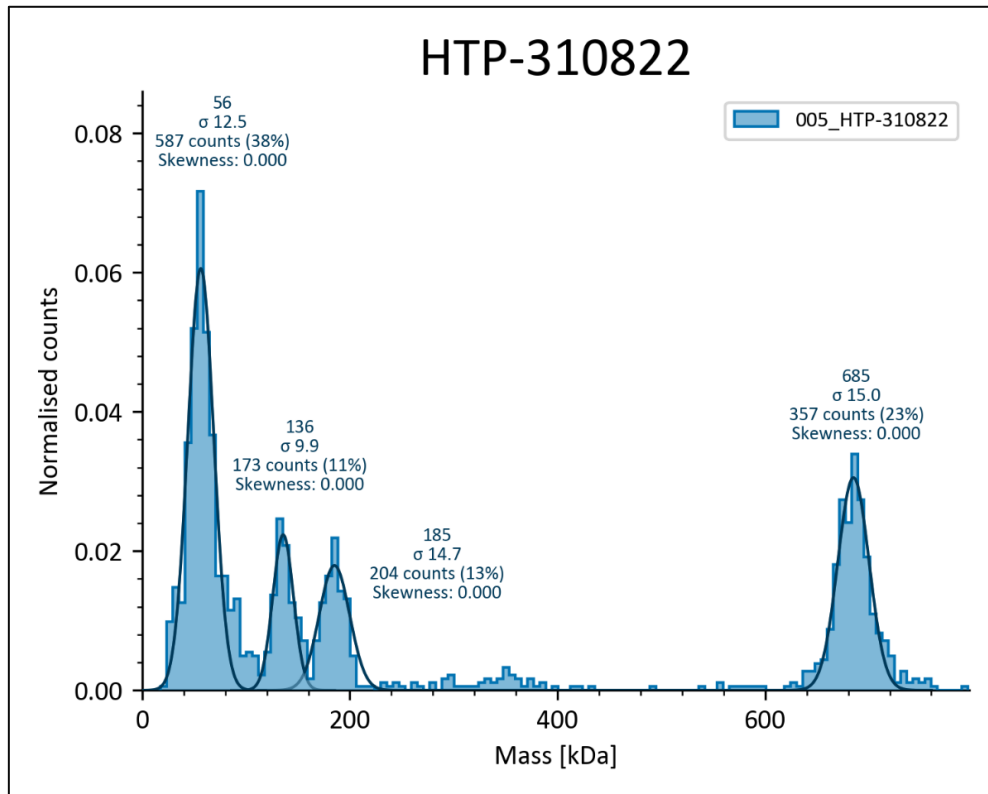
**Figure 1.2-6.** Mesure SEC-SLS de l'échantillon HTP à pH 7,5 et 150 mM NaCl. Le tableau indique les mesures effectuées sur le pic encadré en rouge. Les mesures sont : le poids moléculaire (MW), le temps de rétention (RV), la part du signal associée au pic (Recovery), la part de l'échantillon dans le pic (Frac. of sample), et la concentration dans la fraction du pic (Peak measured conc.). Cette mesure permet de conclure que l'intégralité du pic est peuplé exclusivement par le complexe HTP de taille de 171 364 g/mol à une concentration de 0,29 mg/ml.

### Photométrie de masse (REFEYN)

La photométrie de masse est une méthode qui permet d'estimer la masse d'un objet à la suite de la diffusion de la lumière lors du contact de l'objet avec une surface en verre. Cette méthode permet de travailler en évènement unique avec une résolution de 20kDa.

Les données de mass-photométrie proviennent d'un échantillon vieillissant ayant subi un cycle de congélation-décongélation. Elles montrent l'existence de 5 populations dans l'échantillon. Du plus au moins majoritaire nous retrouvons un pic à ~56kDa (38 %) qui peut être associé au poids moléculaire d'un Fab, un à ~685kDa (23 %) attribuable à des agrégats, un pic à ~185kDa (13 %) attribuable au complexe HTP, et un à ~136kDa (11 %) attribuable à un état partiellement assemblé du complexe dans lequel un des Fab manque. Un dernier pic à peine visible aux environs des 380kDa pourrait correspondre à un dimère du complexe HTP, qui sera

ci-après appelé doublet. Ces observations doivent être mises en parallèle avec l'affinité connue entre les anticorps et HER2 (1,43 nM pour le Trastuzumab et 1,92 nM pour le pertuzumab par test elisa) (Fu et al. 2014) puisque la photométrie de masse est faite à 10 nM ce qui est proche du kD.

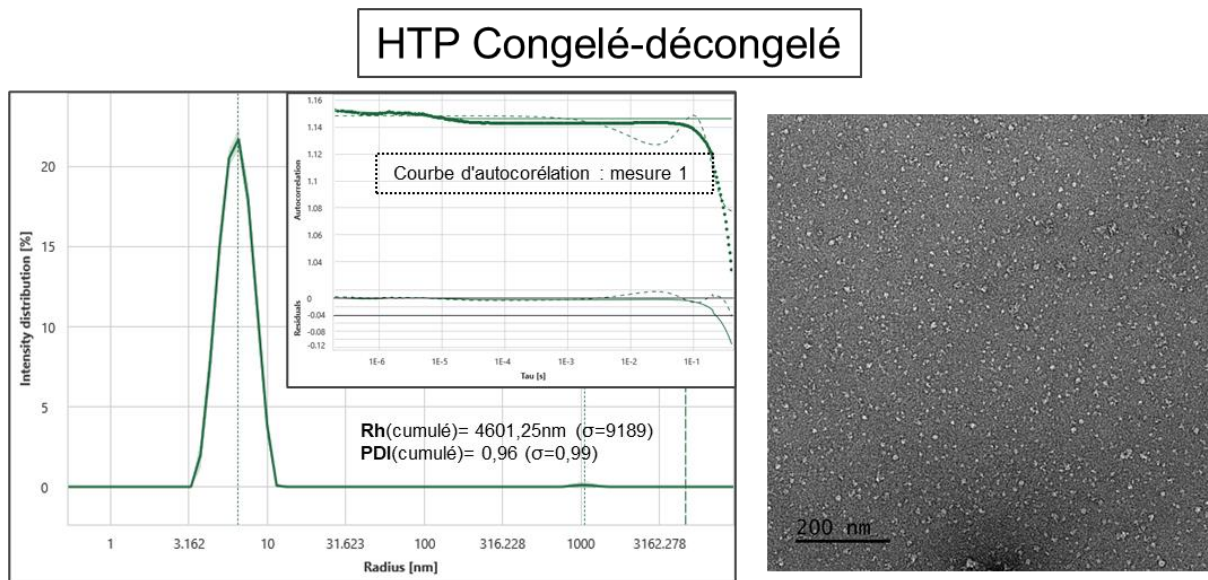
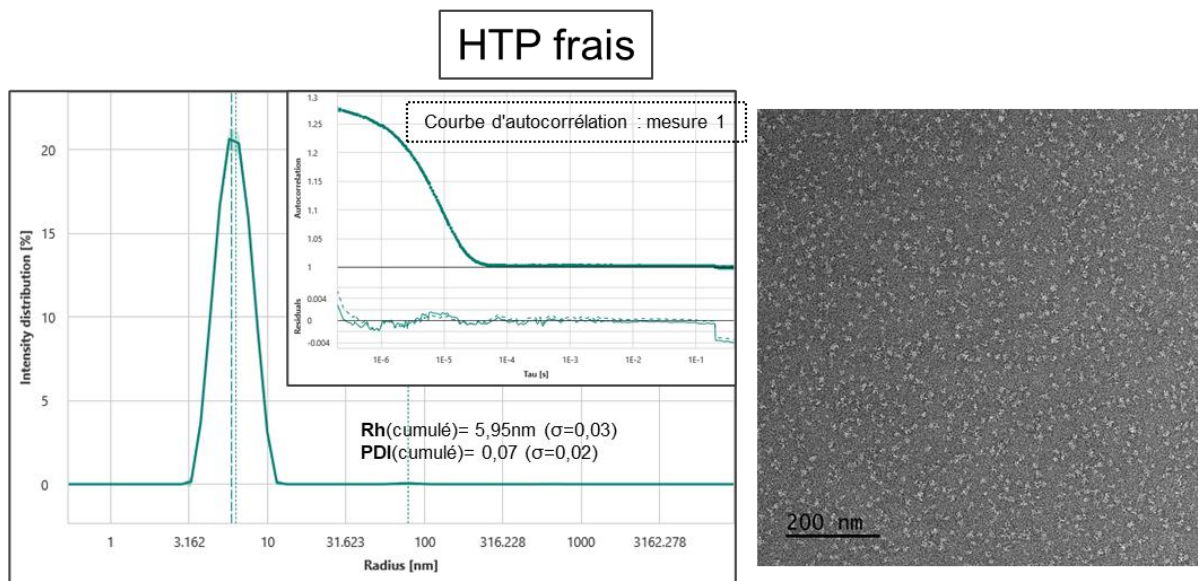


**Figure 1.2-7.** Résultat des mesures par photométrie de masse. Pour chaque pic est attribué une valeur de masse moléculaire en kDa, un écart type ( $\sigma$ ), un nombre de coups mesurés et son pourcentage par rapport à l'ensemble des mesures.

#### 2.1.2.4 Test de stabilité en congélation rapide

Dans la continuité, intéressons-nous aux mesures de DLS et aux observations en coloration négative ci-après (figure 1.2-8). Les résultats présentent une comparaison entre un échantillon frais et un échantillon ayant subi une étape de congélation rapide. Dans les deux cas on observe une variation significative de la taille des objets en solution. Bien que la courbe de distribution de taille semble indiquer le même rayon hydrodynamique, la courbe d'autocorrélation DLS montre un signal fortement perturbé après congélation, indiquant la présence d'objets de grosse taille, probablement des agrégats. De même les valeurs de rayon cumulés et d'indice de polydispersité plafonnent pour l'échantillon congelé alors qu'elles sont à 5,95 nm ( $\pm 0,03$ ) pour un PDI de 0,07 ( $\pm 0,02$ ) pour l'échantillon frais. En coloration négative l'aspect des objets est différent avant et après la congélation. Après congélation l'échantillon change d'aspect et paraît plus petit. Des zones d'agrégats sont aussi visibles à différents endroits de la grille. Ces derniers résultats associés à ceux obtenus par photométrie de masse montrent que l'échantillon est sensible au vieillissement et qu'il supporte mal la congélation.





**Figure 2.2-9.** Test de stabilité sur le complexe HTP. Mesure DLS et observation par coloration négative sur un échantillon frais (panel du haut) contre un échantillon ayant subi une étape de congélation rapide (panel du bas).

Bien qu'un passage sur chromatographie d'exclusion permettrait d'envisager d'utiliser un échantillon congelé, il faudra s'attendre à une grosse perte et peut-être un échantillon moins homogène pour la suite. Les grilles de cryo-EM sont donc préparées le jour de l'assemblage et du passage sur FPLC du complexe.

### 2.1.3 Conclusion et discussion: un échantillon stable et homogène en solution

Les analyses précédentes nous donnent une carte détaillée du complexe HTP et de son comportement face à différents stress. Nous avons présenté un large panel de contrôle qualité de l'échantillon à la fois adaptatif et robuste et avons une vue exhaustive de l'état de l'échantillon en solution.

L'évolution des outils de biophysiques repousse les limites associées à la quantité de matériel nécessaire pour faire une mesure. En ce qui concerne la DSF et la DLS les appareils de la gamme Nanotemper utilisent des capillaires qui ne consomment que très peu de volume d'échantillon (<10  $\mu$ l) et la sensibilité des détecteurs permet de travailler à des concentrations relativement basses. Leur prise en main est très simple ce qui permet d'implémenter ces mesures en routine dans le processus d'optimisation de production et de purification d'échantillon. Cependant les mesures en capillaires sont plus sensibles aux contaminations artéfactuelles tels que les poussières ou les traces sur le verre ou le support de mesures.

Concernant les injections en FPLC, ce n'est pas une méthode infaillible et optimale pour la validation de la formation du complexe. Le protocole le plus robuste consiste à injecter successivement sHER2 seul, puis sHER2 avec un des anticorps, et enfin sHER2 avec les deux partenaires. Par cette approche le temps de rétention est réduit ce qui est le signe d'un changement du rayon de Stokes de l'échantillon, souvent assimilable à une augmentation de sa taille ou de sa forme. Mais cela nécessite de travailler avec un échantillon suffisamment concentré pour réussir à faire plusieurs injections successives. Or, la concentration est souvent un élément limitant. HTP est un cas idéal et nous verrons dans le chapitre suivant que d'autres techniques biophysiques sont plus adaptées pour des tests d'interactions en particulier lorsqu'il s'agit d'anticorps monoclonaux, et spécifiquement de Fabs, qui peuvent être particulièrement coûteux et en quantités limitées.

La photométrie de masse révèle son intérêt dans sa capacité à pouvoir attribuer distinctement chaque événement et ainsi donner un spectre exhaustif des différentes populations qui composent l'échantillon. Un autre point fort de cette approche est la faible quantité de protéines utilisée. Mais la grande sensibilité de la technique est aussi sa faiblesse. Le nombre d'événements étant directement proportionnel à la concentration de l'échantillon déposé sur le capteur, il faut donc travailler à basse concentration pour une résolution optimale, ce qui peut être limité par les affinités des différents partenaires au sein du complexe.

Par souci de clarté, l'ensemble des techniques de caractérisation de l'échantillon sont incluses dans la partie pré-microscope. Cependant il est envisageable d'en réaliser certaines après un premier passage par un microscope. Cependant les données générées par la cryo-EM sont d'autant plus lourdes et chronophages à traiter que l'échantillon sera de mauvaise qualité. Nous verrons ci-après que les étapes vers la collecte introduisent leurs propres incertitudes qui sont parfois difficiles à comprendre.



## 2.2 In-microscope : Préparation de grilles et approche optimisée pour un screening de grilles efficient. La quête de la grille ultime.

À présent que nous savons que le complexe fraîchement préparé est stable et parfaitement soluble dans son tampon standard intéressons-nous à la partie dans le microscope. Comme c'est souvent le cas en cryo-EM nous avons été confrontés au problème de la distribution des particules sur la grille. Nous avons dans un premier temps reproduit les conditions de préparation des grilles cryo-EM comme décrit dans la publication de Hao et al. (Hao et al. 2019) mais il s'avère que nous avons eu de sérieuses difficultés à obtenir de façon reproductible des particules bien distribuées dans la glace. Une quête a donc démarré pour optimiser la préparation des grilles. Toutes les grilles sont achetées auprès de Quantifoil. Les grilles UTC-Quantifoil sont des grilles quantifoil classiques 1.2/1.3 couverte d'un film de carbone (graphène) très fin de 2 nm. Les Ultrafoil sont des grilles intégralement en Or de 0.6/1.

### 2.2.1 Matériel et méthode : la préparation des grilles et le screening sur le microscope

#### 2.2.1.1 Préparation des grilles

Immédiatement avant leur préparation, les grilles ont subi une décharge par plasma avec l'appareil "GloQube plus" ou le système "ELMO glow discharge". Sauf pour les tests effectués sur les quantifoils, le protocole habituel pour les grilles Quantifoil, Ultrafoil, UTC Quantifoil et C-flat est de 10-20 secondes à 12 mA sous vide. Toutes les grilles ont été préparées par "congélation plongeante" dans de l'éthane liquide à l'aide du système Vitrobot Mark IV (Thermo Fisher Scientific, États-Unis). Les paramètres de séchage ont été ajustés pour chaque session et dépendent du vitrobot utilisé. En général, le temps de séchage (blotting time) était d'environ 4 secondes, la force de séchage (blotting force) était de 0 et il n'y avait pas de temps d'attente avant le séchage. L'échantillon déposé était de 2 à 3  $\mu$ l avec une OD 280 nm comprise entre 0,1 et 0,3, dans une chambre d'humidité maintenue à 100 % pour toutes les préparations de grilles. Des tests sur la température de préparation de l'échantillon ont montré que celle-ci n'avait pas d'impact sur la distribution de l'échantillon sur la grille, ce qui est cohérent avec les mesures DSF et DLS, et les grilles ont donc été préparées à 15°C. Pendant la procédure, l'échantillon a été conservé dans la glace. D'un côté sur des grilles quantifoil classique nous avons fait varier des paramètres tels que la composition du tampon, la procédure de traitement par plasma, et les conditions de vitrification du vitrobot. D'un autre côté, nous avons testé une grande variété de grilles. Toutes ces expériences sont résumées dans le tableau 2.3.1.3-1.

#### 2.2.1.2 Visualisation sur le microscope à transmission

Le screening a été fait sur deux microscopes. Une partie des tests a été effectuée à l'I2BC sur un Tecnai 200kV équipé d'une caméra K2 Summit (Gatan) et d'un chargement de grille unique. Les autres tests ont été réalisés à la plateforme nano Imaging de l'institut Pasteur sur un Glacios 200kV, équipé d'une caméra Falcon 3 ou Falcon 4i sans filtre à énergie. Le logiciel EPU

a été utilisé pour le contrôle du microscope et la visualisation des images. Les caractéristiques des microscopes sont reprises dans le tableau 3.1-1.

### 2.2.1.3 Stratégie de screening de condition de préparation de grille

Nous avons d'abord tenté de reproduire le protocole publié en 2019 par Hao et al. (Hao et al. 2019), dans lequel le complexe était préparé sur des grilles quantifoil classique. Nous nous sommes ainsi aperçus que les particules étaient dans la majeure partie des grilles mal distribuées. Nous avons donc exploré plusieurs axes pour tenter d'élucider le mystère à l'origine du comportement des particules sur ce type de grille. Nous avons ainsi observé l'effet de différents paramètres physiques de la grille et biochimiques de l'échantillon sur les grilles quantifoil. Puis, d'un autre côté nous avons observé un effet du matériau de la grille sur la distribution des particules. Chaque condition de screening est préparée en duplicata ou triplicata. L'observation est systématiquement poussée jusqu'à la visualisation des particules dans les trous. La quantifoil classique a été testée et préparée plusieurs fois dans différents environnements.

Pour les tests de pH le complexe est assemblé comme décrit dans le paragraphe 2.1.3 puis injecté et élué sur une colonne S200 préalablement équilibrée dans le tampon correspondant.

**Post-traitement des grilles au MgAc.** Elles sont dans un premier temps passées dans la chambre à plasma puis trempées dans une solution de MgAc 5 mM à deux reprises avant d'être séchées avec du papier Wattman avant une utilisation classique dans le vitrobot. Le protocole doit rendre les grilles hydrophiles et chargées positivement.

**Traitements par bain d'acétone, éthanol et SDS.** Les grilles sont trempées dans les solutions pendant 2 à 3 heures. Ensuite elles sont séchées par contact sur du papier Wattman puis brièvement réintroduites dans la solution avant d'être rincées 2 fois dans l'eau distillée, par des cycles de rinçages/séchages sur du papier Wattman. Les grilles sont ensuite montées sur le vitrobot pour une utilisation classique.

**Les paramètres de vitrobot .** Les grilles sont préparées dans différentes conditions de température, de temps et de force de blotting, ainsi que l'introduction d'un temps d'attente avant le blotting. Trois vitrobot différents ont été testés (I2BC, Pasteur, Sanofi).

**Le traitement par plasma.** Les paramètres ont été modifiés sur l'appareil GloQube plus system. Toutes les combinaisons entre une intensité de 5, 10, 25 et 50 mA et un temps de 10, 20 et 40 secondes ont été testées.

**La concentration de l'échantillon** a été testée à une DO de 0,1 et 0,3. Une collecte a été faite avec une DO de 0,3 (0.12 mg/ml ou 0.73  $\mu$ M).

**Concentration de sel.** L'échantillon est élué dans un tampon à 150 mM NaCl puis dilué à une DO de 0,15 en augmentant ou diminuant la quantité de sel à 30 ou 300 mM. Les trois

concentrations de sels sont testées sur le même échantillon le même jour.

**Les tests de détergent.** Pour étudier leur effet les détergents sont ajoutés dans une aliquote de protéines fraîchement éluée quelques instants avant la vitrification. Le détergent est préalablement dilué dans le tampon de la protéine à une concentration de 5X puis 2 µl de détergent est ajouté dans 10 µl de solution de protéine. La concentration finale pour chacun d'entre eux est de 0,2 mM pour le SDS, 0,125 % pour le Tween20, et 0.01 % pour le TritonX.

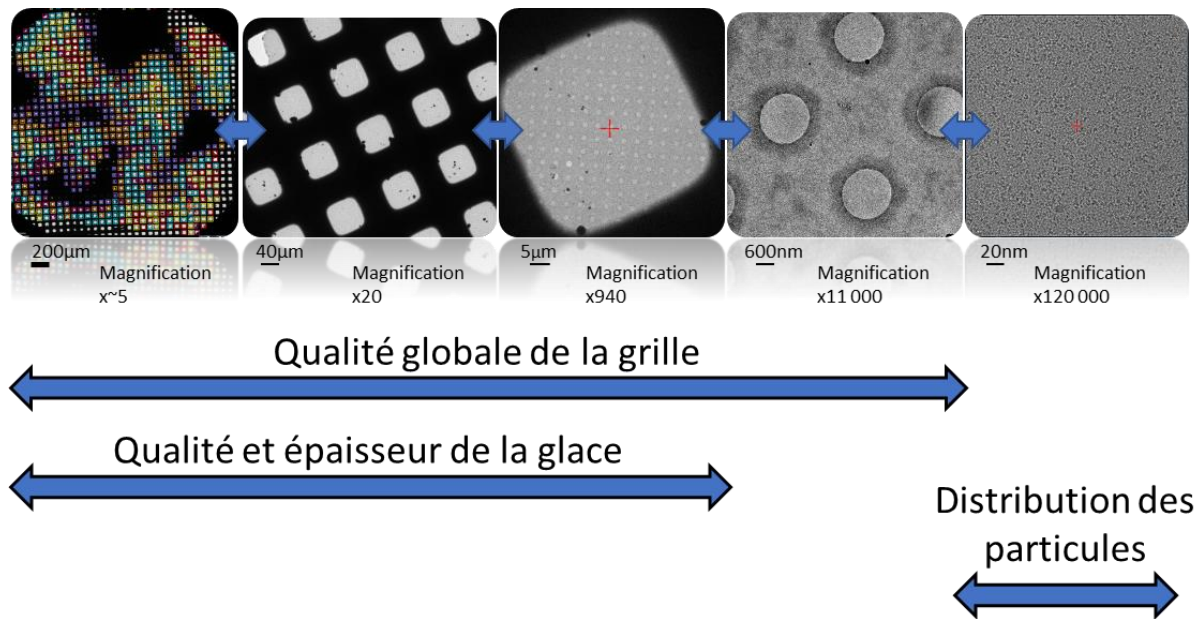
Les grilles sont toutes préparées par vitrification plongeante avec un vitrobot mark IV.

Résumé des paramètres testés			
Paramètres		Part. dist.	
<b>Grid treatment</b>	5 mM MgAc post-plasma, 10 % Acetone, 10 % EtOH, 10 % SDS.	Fail	Grids physics
<b>Blotting T°</b>	4°C, 15°C, 22°C		
<b>Blotting parameters</b>	Blotting times, blotting force, waiting time		
<b>Plasma cleaner</b>	5-50mA 0-40sec		
<b>Concentration</b>	DO~0,3	Partial success	Biochemical condition
<b>pH</b>	4,3 – 7,5 – 9	Fail	
<b>Salt</b>	NaCl 30/150/300 mM		
<b>Buffer detergent</b>	SDS 0,2 mM & 0,1 %, Tween 20 0,125 %, TritonX 0,01 %		
<b>Grids</b>	Quantifoil (Cu, Au, C-Flat) & Lacey	Fail	Grids Nature
	UTC Quantifoil	Success	
	UltrAufoil	Partial success	

**Tableau 2.1.3-1.** Bilan des paramètres testés lors de l'optimisation de la préparation des grilles.

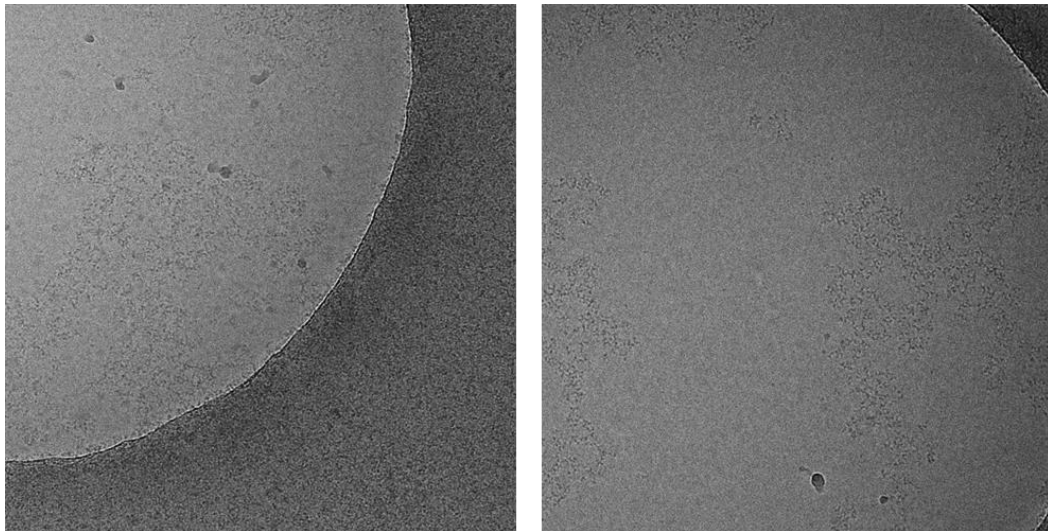
### 2.2.2 Résultats : un effet grille dépendant

L'échantillon fraîchement préparé est systématiquement caractérisé par DLS et nano-DSF, puis déposé sur grilles et vitrifié dans les diverses conditions décrites dans le tableau 2.1.3-1. Pour rappel, HTP est une particule de 170kDa dont les dimensions sont de 15 nm par 7 nm. C'est donc un objet de bien trop petite taille pour distinguer l'état du complexe au moment de l'observation des images brutes. Au moment de l'observation de la grille sur le microscope, nous jugeons seulement la qualité de la glace et de la grille, la distribution et la densité des particules. Le premier jugement qualitatif est apporté par la classification 2D réalisée après la collecte. Chaque critère qualité est jugé à différent grossissement (figure 2.2.2-1).



**Figure 2.2.2-1.** Représentation schématique des différents grossissements et des critères qualité associés.

Dans les conditions standard avec une grille Quantifoil classique, les particules sont systématiquement mal distribuées, indépendamment de l'épaisseur de la glace (figure 2.2.2-2). De même aucune des tentatives de traitement chimique des grilles avant ou après le passage par traitement plasma n'a changé le comportement des particules dans les trous. Enfin, la variation des paramètres du vitrobot tels que la force, le temps, la température et l'humidité n'a pas amélioré la distribution des particules.

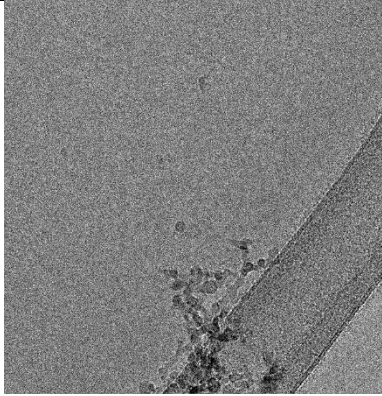
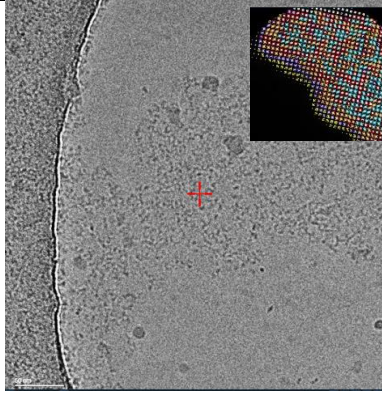
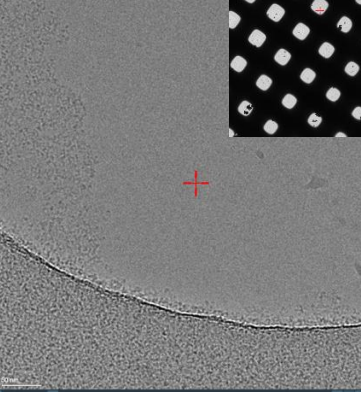
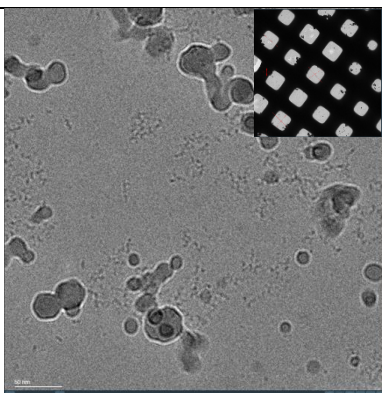
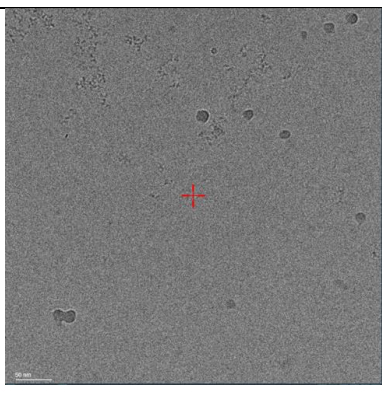
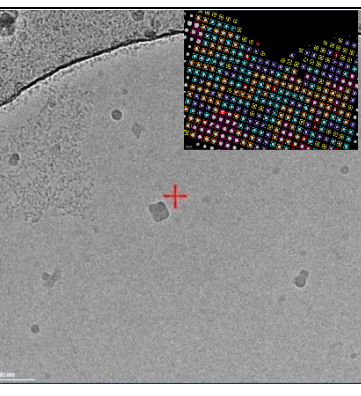
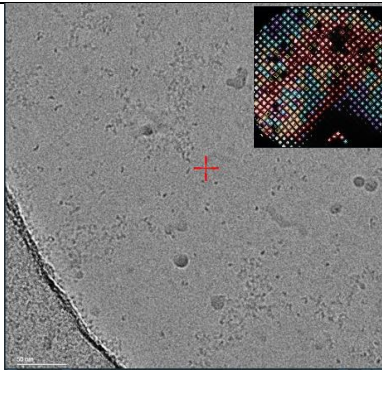
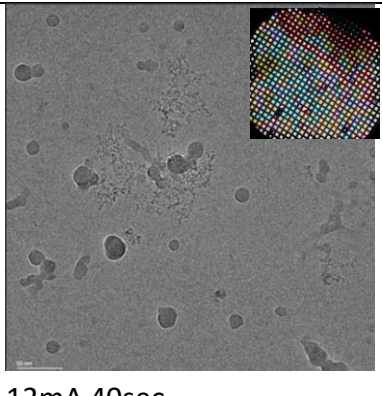
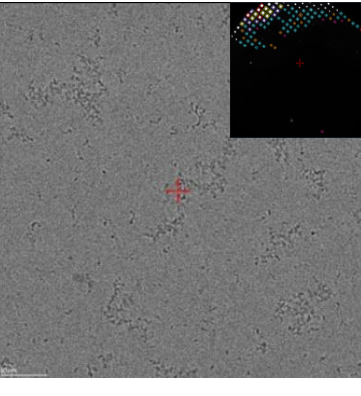


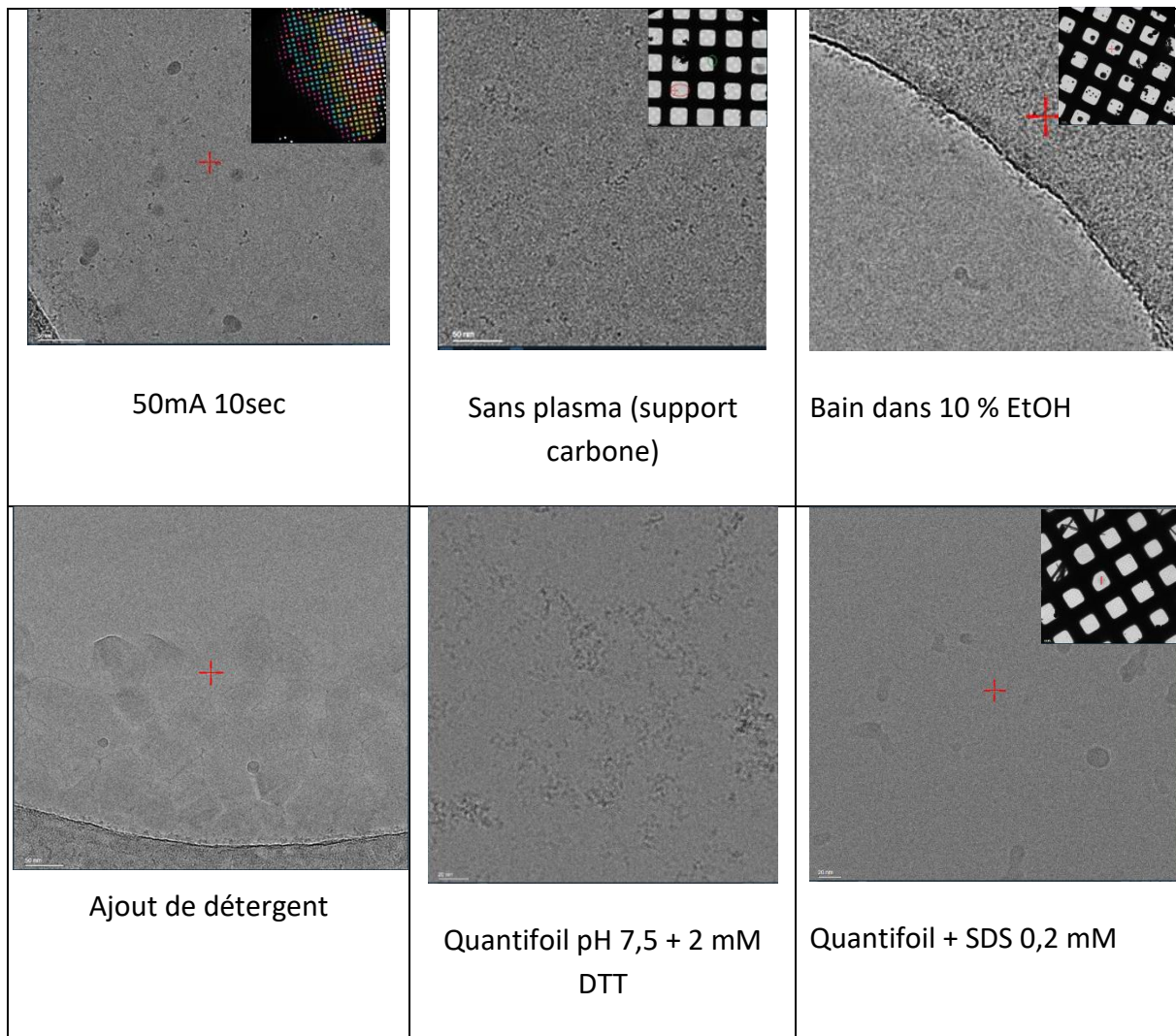
**Figure 2.2.2-2.** Exemple de la distribution des particules sur grilles quantifoil classiques. Images collectées sur un Tecnai FEI 200kV équipé d'une caméra K2 Gatan.

Les changements des paramètres de traitement plasma n'ont pas changé la distribution des particules. Ces derniers ont plutôt révélé une tendance à une glace d'autant plus homogène que le temps et l'intensité du courant sont élevés. En revanche, sur les grilles avec un revêtement continu de carbone, il y a plus de dégât sur le support carbone à de fortes intensités. Inversement l'absence de traitement plasma conduit sans surprise à des grilles de très mauvaise qualité. La glace n'est pas homogène ni sur la grille, ni au sein des mailles et de nombreuses taches noires (similaire à l'impact d'une goutte d'encre ou parfois appelée "pizza") sont visibles sur la grille. Les incubations préliminaires dans des solutions contenant 10 % d'éthanol ou 10 % d'acétone mènent aux mêmes observations, avec des taches noires et une mauvaise répartition de glace vitreuse, qui sont les signes d'une surface qui est restée hydrophobe. Enfin notons l'absence de particules collées sur le support carbone qui est normalement saturé lors des traitements plasma standard (tableau 2.2.2-1). Le paramétrage optimal de la machine est situé autour de 12mA pendant 20 secondes. Ces résultats indiquent que ce ne sont pas les charges à la surface de la grille qui impactent la distribution des particules dans les trous.

Aucun des tests sur la physique de la grille n'a permis de résoudre la mauvaise distribution des particules.



		
<p>Lacey grids</p>	<p>Quantifoil pH7,5</p>	<p>Quantifoil pH7,5 + 5 mM MgAc</p>
		
<p>Quantifoil pH 9</p>	<p>Quantifoil pH 9 + MgAc</p>	<p>Quantifoil Ac pH 4</p>
		
<p>5mA 10sec</p>	<p>12mA 40sec</p>	<p>5mA 40sec</p>



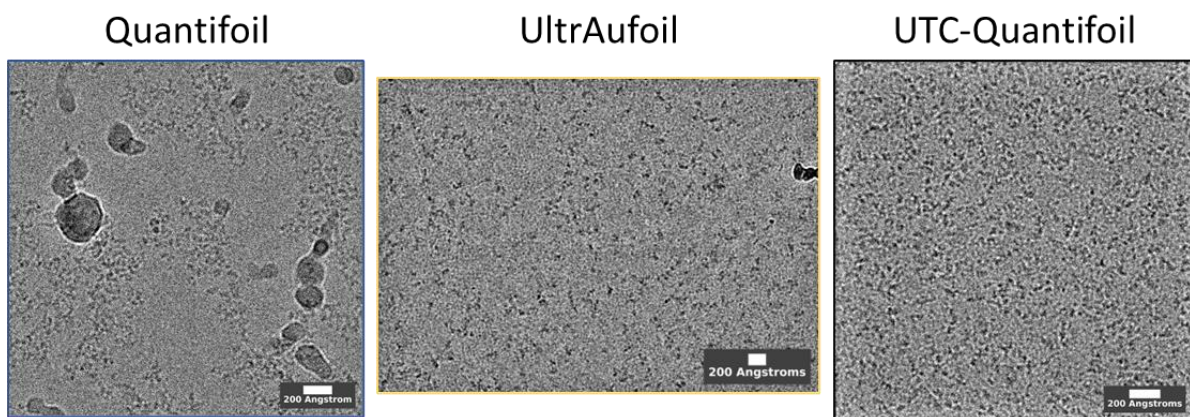
**Tableau 2.2.2-1.** Quelques résultats des principales conditions testées. Lorsque le tampon n'est pas indiqué la grille est préparée dans le tampon standard (HEPES pH7,5 150 mM NaCl).

De même les différents tests sur les conditions biochimiques, à savoir le changement de concentration de la protéine, du sel, les changements de pH ou l'ajout de détergent, n'ont pas impacté la distribution des particules sur les grilles. Comme nous l'avons précédemment montré dans la partie pré-microscope, l'échantillon est relativement stable à pH 4,5 et surtout 9. Ces deux pH ont donc été testés en concomitance avec différentes combinaisons de concentration de NaCl (30 mM, 150 mM et 300 mM), mais cela n'a pas amélioré la distribution des particules sur les grilles quantifoil.

L'ajout de détergent dans l'échantillon n'a pas résolu le problème. Au contraire les détergents ont souvent entraîné l'apparition d'artefacts. Les premières observations n'ont pas conduit à une amélioration, ils ont donc été mis de côté.

L'augmentation de la concentration de l'échantillon à une DO de 0,3 sur une grille quantifoil a permis d'augmenter la densité de particules visibles dans les trous et ainsi faire une collecte, bien qu'elles soient toujours mal distribuées.

Enfin concernant le changement du type de grille, les grilles quantifoil en cuivre, quantifoil en or avec support carbone, les C-Flat et les laceys n'ont pas amélioré la distribution des particules dans les trous. En revanche les grilles Ultraufoil intégralement en or conduisent de manière systématique à des particules bien distribuées dans les trous, mais elles requièrent de travailler à une  $DO \geq 0,3$  sous peine d'avoir une trop faible densité de particules. De même les grilles de type UTC-quantifoil, grilles avec un très fin (2 nm) film continu de carbone, conduisent à une réussite systématique avec une excellente densité de particules même à faible concentration (Figure 2.2-2). Notons que les UTC-quantifoil sont très fragiles et à plusieurs reprises le film de carbone s'est brisé. Il faut donc être vigilant, bien vérifier l'intégrité de la grille, et préparer la grille en duplicata ou triplicata.



**Figure 2.2.2-2.** Images obtenues avec les différents types de grille. Collecté sur Glacios 200kV.

### 2.2.3 Conclusion et discussion : la stratégie de screening de grilles pour une collecte dans des conditions optimales

En conclusion, la variation des conditions biochimiques, tout en maintenant un état stable des particules en solution, n'a pas permis d'améliorer la distribution des particules sur les grilles quantifoils. Seul le changement du type de grilles, en utilisant des grilles entièrement constituées d'or nommées UltrAuroil ou celles recouvertes d'un film de carbone très fin (2 nm) nommées UTC-Quantifoil, a radicalement résolu le problème. Habituellement ce comportement de l'échantillon est attribué à une perturbation liée à l'interface air-eau ou bien à une conséquence de la vitrification. Cependant le fait de ne pas l'avoir observé sur les grilles en or fragilise cette hypothèse. Un effet lié à la présence du cuivre peut aussi être écarté car des grilles quantifoil avec un maillage en or ont été testées et ont donné les mêmes résultats que les quantifoil classiques.

L'hypothèse d'une perturbation liée à la vitrification est envisageable. Il est montré que la vitesse de congélation impacte directement les dommages sur les particules (Klebl et al. 2020). Les différentes populations observées sur le Refeyn, ainsi que les comparaisons entre un échantillon frais et un échantillon ayant subi un cycle de congélation par coloration négative ou DLS sont des éléments à la charge d'un effet lié à la congélation. La conductivité thermique de l'or ( $318 \text{ W.m}^{-1}.\text{K}^{-1}$ ) est légèrement inférieure à celle du cuivre ( $401 \text{ W.m}^{-1}.\text{K}^{-1}$ ), mais drastiquement supérieure à celle du carbone ( $155 \text{ W.m}^{-1}.\text{K}^{-1}$ ). Ainsi nous pouvons supposer une meilleure vitrification sur le support en or que celui en carbone. Cette différence est une piste pour expliquer les améliorations observées sur la grille en or.

A première vue cela pourrait paraître contradictoire étant donné le succès des grilles UTC-quantifoil, entièrement composées de carbone. Mais, l'introduction d'un film continu de graphène change drastiquement l'environnement. Les protéines ont une grande affinité pour le carbone et vont donc se "coller" à celui-ci, ce qui pourrait avoir un effet protecteur lors de la vitrification. De plus, le graphène a une meilleure conductivité thermique que la glace, il va donc peut être avoir un effet potentiel lors de la vitrification. Enfin l'un des grands intérêts du film continu de graphène, est son effet protecteur face aux radiations du faisceau en dissipant la chaleur et réduisant la sublimation de surface (Kato et al. 2021)

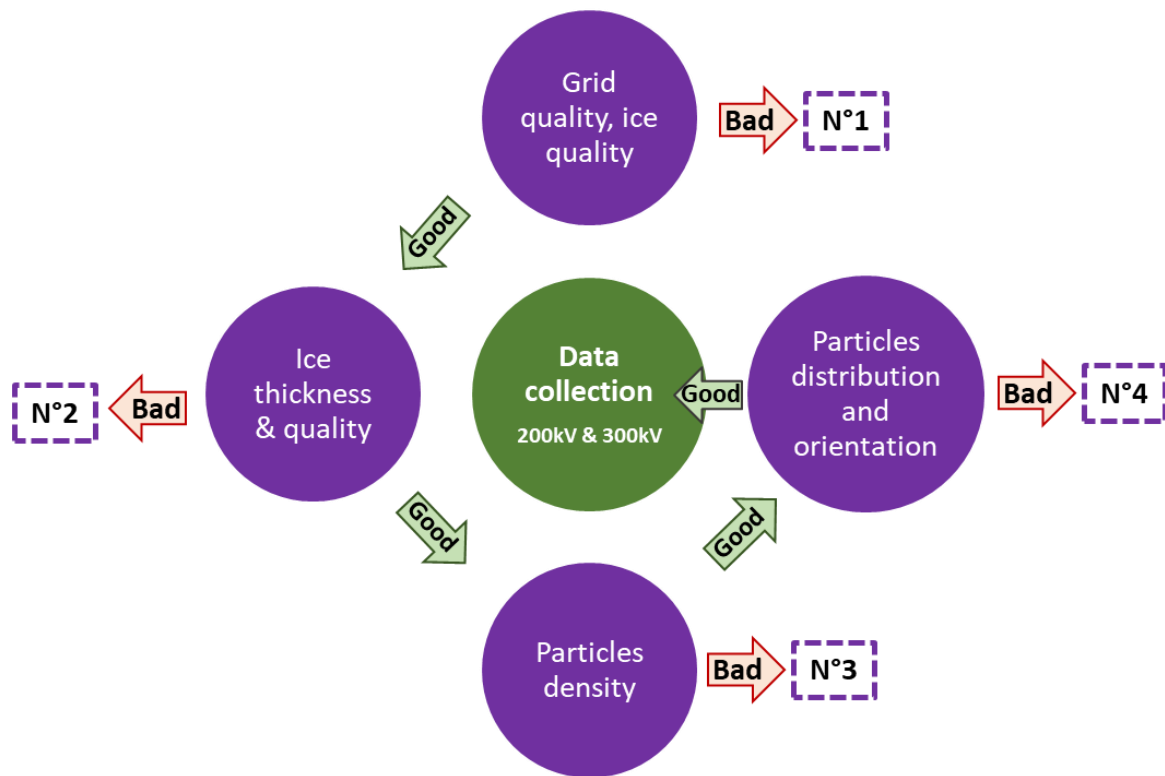
Seul le vitrobot (vitrification plongeante) a été testé, et peut-être qu'une autre technique de préparation des grilles permettrait de mieux maîtriser le processus de vitrification et donc le comportement des particules sur la grille (Klebl et al. 2020; Weissenberger et al. 2021).

Bien qu'un large spectre de conditions ait été testé et que deux solutions aient été identifiées, des interrogations concernant le comportement des particules persistent. Certaines conditions manquent, parmi celles-ci le changement de sel pour un sel de magnésium ou de potassium par exemple. De même l'ajout d'arginine à 100mM est connu pour ses effets stabilisants des protéines en solution (Tsumoto et al. 2004) (Fan et al. 2021) et serait donc une piste de choix à explorer. Néanmoins il faut prendre soin de vérifier préalablement que cela

ne perturbe pas la fixation des Fabs par DLS, nano-DSF et photométrie de masse par exemple.

Il est possible que la mauvaise distribution des particules sur les grilles quantifoil en cuivre classique provienne du contact des protéines avec l'interface air-eau. Dans ce cas l'utilisation de détergent en faible concentration est souvent bénéfique à la distribution et à l'orientation des particules (Baker et al. 2014; Kampjut et al. 2021). Les tests succincts avec les trois détergents pourraient être complétés par l'utilisation de kit plus exhaustif. En effet nous avons testé trois détergents donc un ionique (SDS) et deux non ioniques (Triton et Tween). Peut-être que d'autres détergents du même type ou zwitterionique aurait pu permettre d'améliorer les grilles. Cependant ces tests auraient requis de longues nouvelles sessions de screening de grilles, gourmandes en temps et en matériel. De plus, aucune tendance à l'amélioration n'a pu être dégagée des premiers essais et il en résulte principalement une baisse du contraste de l'image et l'observation d'artefacts liés au détergent. Il nous a donc paru plus direct et optimal de tester en priorité les autres familles de grilles qui se sont avérées particulièrement efficaces pour résoudre le problème de distribution des protéines sur la grille.

Ces essais nous ont amené à une réflexion approfondie sur l'approche optimale de la préparation de grille. Ils aboutissent à la concrétisation d'un protocole standardisé du contrôle des grilles, ainsi qu'un guide synthétique hiérarchisé du choix le plus adapté lors du screening des grilles pour ce type de petit échantillon (figure 1.6-1).



### N°1 - Grid overall quality

- Change blotting parameters (Blotting time, force, humidity, temperature, deposit volume)
- Change grid type
- Change glow discharge parameters

### N°2 - Ice thickness

- Change blotting parameters (blotting time + force, sample  $\mu$ l)
- Change grids holes size
- Change deposit methods (*D'Imprima and Kühlbrandt 2021*)

### N°3 - Particles density

- Try continuous carbon grids or continuous graphene grids
- Control the concentration (*Vinothkumar, K., & Henderson, R., 2016*)

Number of particles in projection/ $\mu$ m<sup>2</sup> in 800 Å thick ice film (separation)

M.W.	Concentration				
	10mg/ml	2mg/ml	0.5mg/ml	0.1mg/ml	20 $\mu$ g/ml
10 kD	48000 (45Å)	10000 (100Å)	2500 (200Å)	500 (450 Å)	100 (1000 Å)
50 kD	10000 (100Å)	2000 (220Å)	500 (400Å)	100 (1000Å)	20 (0.2 $\mu$ m)
250kD	2000 (220Å)	400 (500 Å)	100 (1000 Å)	20 (0.2 $\mu$ m)	4 (0.5 $\mu$ m)
1 MD	500 (400Å)	100 (1000Å)	25 (0.2 $\mu$ m)	5 (0.4 $\mu$ m)	1 (1 $\mu$ m)
5 MD	100 (1000Å)	20 (0.2 $\mu$ m)	5 (0.4 $\mu$ m)	1 (1 $\mu$ m)	0.2 (2.2 $\mu$ m)
25 MD	20 (0.2 $\mu$ m)	4 (0.5 $\mu$ m)	1 (1 $\mu$ m)	0.2 (2.2 $\mu$ m)	0.04 (5 $\mu$ m)

- Detergent screening
- Vitrification methods (*D'Imprima and Kühlbrandt 2021*)

### N°4 - Particles distribution

1

#### Grid type

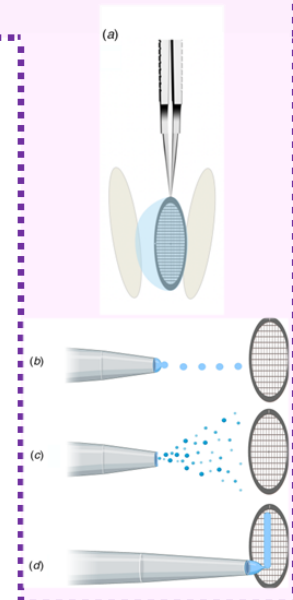
UltrAuroil,  
Continuous  
graphene/carbon  
grids, Quantifoil,  
Lacey...

NO

2

#### Biochemical conditions

Protein  
concentration,  
Salt, pH, detergent



**Figure 1.6-1.** Synthèse et hiérarchisation des points de contrôles majeurs lors de l'étapes de criblage des grilles. Pour chaque problème des solutions sont proposées et hiérarchisées.



## 2.3 Post-microscope : Le traitement des données

La partie in-microscope s'est conclue par des collectes de données sur trois types de grilles différents. Dans cette partie nous allons plonger dans l'analyse de ces données pour en apprendre plus sur l'état de l'échantillon. Nous commencerons par les analyses des grilles Quantifoil et UltraFoil puis la grille UTC-Quantifoil.

### 2.3.1 Matériel et méthode : de la collecte au traitement des données

Pour les parties suivantes six jeux de données ont été collectés et les paramètres principaux liés à chaque collecte sont résumés dans le **tableau 3.1-1**. Un traitement de données exhaustif a été fait sur trois jeux de données, un pour chaque type de grille. Les résultats présentés sur la quantifoil classique et UTC-Quantifoil sont issues de données enregistrées au format eer (caméras Falcon 4i/4). Quant aux résultats obtenus sur l'UltraFoil les films ont été enregistrés au format tiff (caméra K3, avec filtre bioquantum).

Grid type	OD <sub>280nm</sub> [mg/ml] ( $\mu$ M)	Microscope	Data collection parameters						Med. CTF	Ptc. nbrs.	Map. Res.	b-factor
<b>UTC</b> <b>Quan.</b> <b>1.2/1.3 <math>\mu</math>m</b>	0.15 [0,12] (0.73)	Glacios	8787 mic.	240k mag.	Falcon 4	0.58 A/pix	60e/A <sup>2</sup>	3.2Å	711 008	3.1Å	130.2	
UTC 1.2/1.3 $\mu$ m	0.34 [0,27] (1,6)	Glacios	1000 mic.	240k mag.	Falcon 4i	0.58 A/pix	60 e/A <sup>2</sup>	3.7Å	70 071	5.97Å	374.4	
<b>Quantifoil</b> <b>300</b> <b>1.2/1.3 <math>\mu</math>m</b>	0.3 [0,23] (1,45)	Glacios	8870 mic.	240k mag.	Falcon 4i	0.60 A/pix	50e/A <sup>2</sup>	3.5Å & 8.5Å	489 795	<6Å	331	
Quantifoil 300 1.2/1.3 $\mu$ m	0,1 [0,08] (0,48)	Glacios	1448 mic.	120k mag.	falcon 3	1.12 A/pix	60e/A <sup>2</sup>	7.5Å	***	***	***	
<b>UltraFoil</b> <b>300</b> <b>0.6/1 <math>\mu</math>m (coll. 2)</b> <b>*</b>	0.1 [0,08] (0,48)	Titan Krios	4231 mic.	102K mag.	K3/bioquantum	0.86 A/pix	60e/A <sup>2</sup>	3.8Å	740 544	6<Å<7.2	741.3	
UltraFoil 300 R1.2/1.3 $\mu$ m (coll. 1) *	0.1 [0,08] (0,48)	Glacios	1861 mic.	120k mag	Falcon3	1.12 A/pix	60e/A <sup>2</sup>	7.5Å	91 501	7.4Å	662.8	

**Tableau 3.1-1** Résumé des paramètres de collecte et des résultats du traitement des données. Seules les grilles en gras sont référencées dans cet article. La quatrième colonne indique le nombre de micrographies (mic.) restantes après sélection. La cinquième colonne indique le grossissement (mag) utilisé pendant la collecte. La sixième colonne indique la taille des pixels des films RAW. La neuvième colonne indique la valeur médiane approximative de la résolution des micrographies selon l'estimation de CTF et a été déterminée visuellement selon le tracé CTFfitresolution=f(defocus) réalisé dans cryosparc. Le facteur B n'est donné que pour l'affinement global. \* deux jeux de données collectés sur la même grille

#### 2.3.1.1 Quantifoil classique

##### Pre-processing

Les données ont été traitées avec cryosparc V3.3.2. et cryosparc V4.1.2. Deux jeux de données au format eer (un de 5566 films et l'autre de 5958 films) collectés sur le même microscope, la même grille, et au cours de deux sessions différentes mais avec les mêmes paramètres de collecte ont été fusionnés. Ils ont été importés dans cryosparc avec un fractionnement eer de 50 soit une dose de 1 électron par fraction. La taille du pixel vaut 0,6 Å/pix. Les images des films sont alignées avec le "patch motion correction" en utilisant les paramètres standards en



utilisant la pleine résolution et toutes les images soit une dose totale de  $50 \text{ e}^-/\text{\AA}^2$ . L'estimation de CTF est faite avec le "patch CTF estimation" avec les paramètres standards c'est-à-dire une gamme de résolution entre 4 à 25 Ångstrom, et une gamme de défocalisation entre 0,1  $\mu\text{m}$  et 4  $\mu\text{m}$ . En utilisant l'outil de tri "manually curate exposures" les micrographies sont ensuite triées selon trois critères, "total full frame motion distance", "average defocus", et "CTFfit resolution", celles associées à des valeurs extrêmes sont retirées. Il en résulte un total de 8870 micrographies qui sont utilisées pour la suite.

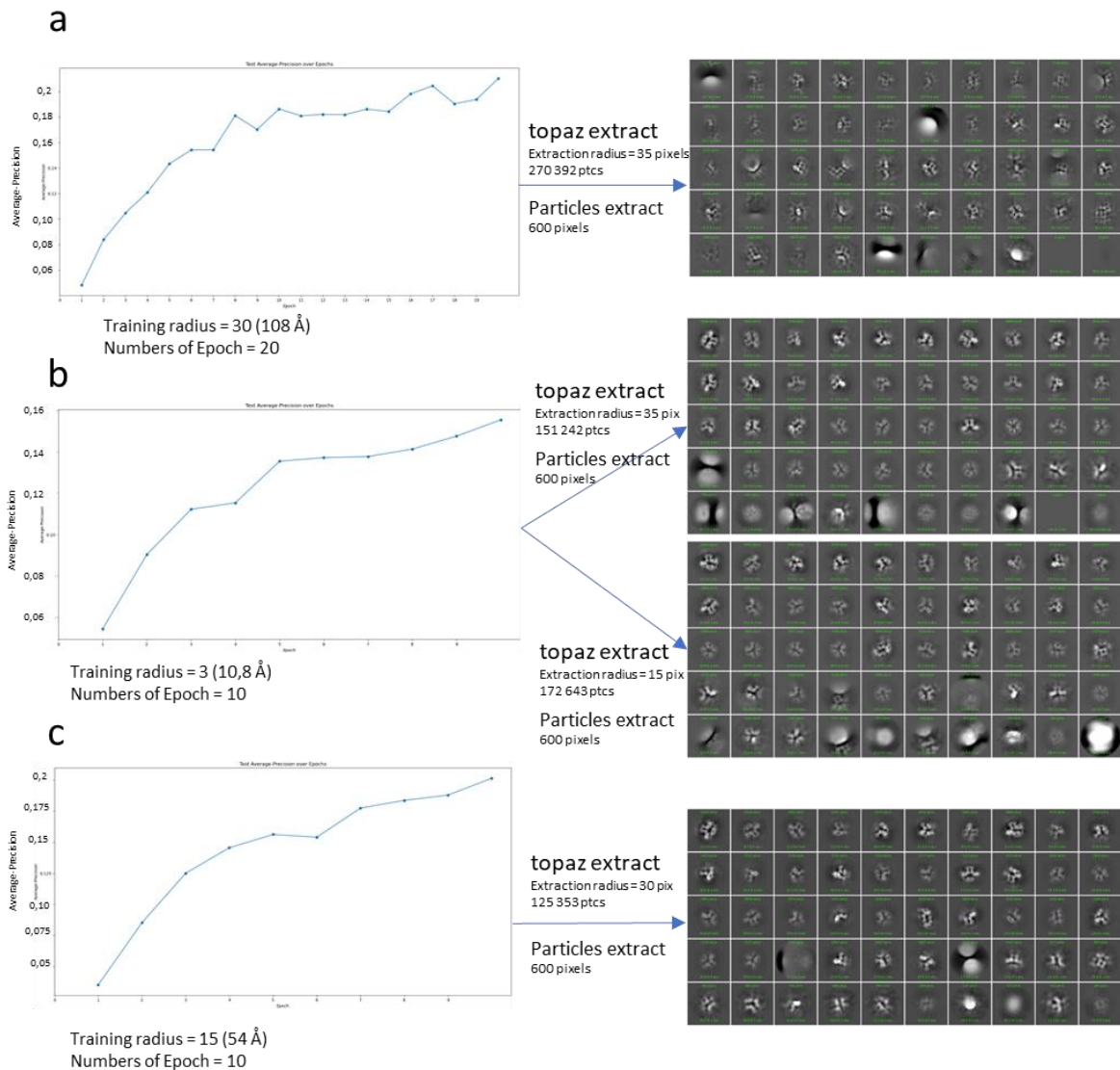
### La sélection (picking) et le tri des particules

Pour commencer, un blob picking est fait sur un sous jeu de 300 micrographies prises au hasard, une taille comprise entre 120 et 250 Å pour une particule de dimension 7 nm par 15 nm, un filtre de résolution à 20 Å et une distance minimale de 1 soit 120 Å. Pour retirer un maximum de faux positifs, les coordonnées obtenues sont triées via le "inspect particles pick" puis les zones sont extraites avec une taille de boîte de 420 Å rééchantillonnée à 108 Å. Les particules sont soumises à une classification 2D avec 50 classes avec les paramètres par défaut de cryoSPARC. Quinze modèles 2D sont utilisés pour un "template picking" sur l'ensemble du jeu de données. Les coordonnées sont inspectées puis 683695 particules sont extraites en utilisant une boîte de 360 Å rééchantillonnée à 180 Å. Les particules issues de 9 classes 2D sont sélectionnées pour créer 3 volumes ab-initio représentatifs des bonnes particules et qui seront utilisés pour les étapes ultérieures de classifications 3D.

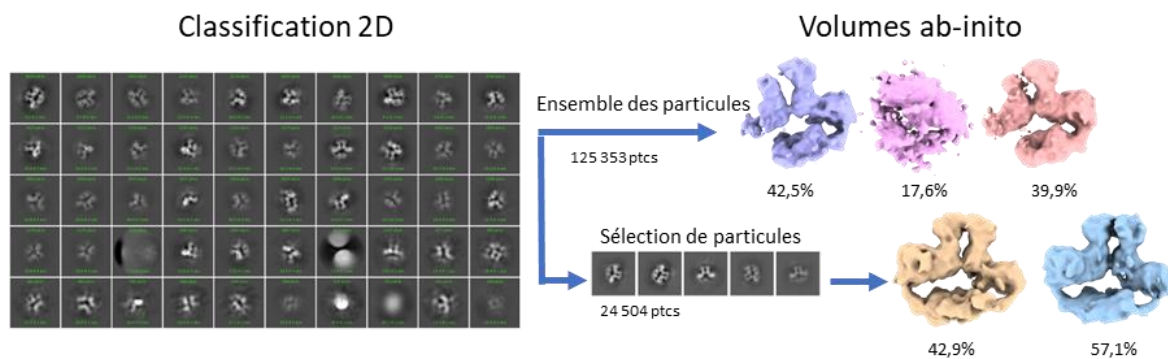
Certaines classes 2D suggérant des formes de "doublet" de HTP (dimère de trimère), nous avons utilisé un logiciel de recherche de particules intelligent nommée topaz (Bepler et al. 2019) afin d'enrichir en particules la population du doublet HTP. À partir de la classification 2D, un modèle montrant le doublet a été utilisé pour un template picking sur 1000 micrographies. Les 79594 particules extraites ont été redimensionné d'un facteur 2 (taille de pixel 1,12 Å) avant d'être triées par deux classifications 2D successives en ne retenant que les classes montrant un doublet. Les 1286 particules restantes et leurs micrographies associées ont été sélectionnées et utilisées pour l'entraînement de topaz (topaz train). L'entraînement a été effectué avec un facteur de rééchantillonnage de 6 (taille effective du pixel de 3,6 Å), un rayon d'entraînement de 15 pixels (54 Å) et un nombre prévu de 10 particules par micrographie. Les autres paramètres ont été conservés par défaut. Ces paramètres ont été optimisées à la suite de nombreux test avec différentes valeurs du rayon d'extraction (figure 2.3.1.1-1).

L'extraction Topaz (topaz extract) a été effectuée sur l'ensemble des micrographies, le rayon à considérer pour l'extraction étant fixé à 30 pixels (108 Å). Durant cette étape l'extraction ne consiste pas à extraire l'image de la particule mais simplement identifier les coordonnées des particules, c'est finalement une étape de "picking".

Les particules sont extraites dans une boîte de 360 Å réduite à 180 Å. Les particules sont classées en trois classes 3D en utilisant l'affinement hétérogène (heterogeneous refinement) et une taille de pixel de 2,5 Å (Voxel 144). Les particules assignées à la classe des doublets ont été combinées avec celles obtenues par template picking (figure 2.3.1.1-2).



**Figure 2.3.1.1-1.** Différentes stratégies topaz à partir du même lot de particules initial. Les paramètres qui impactent le plus la qualité du "picking" sont le training radius et l'extraction radius. Un rayon d'entraînement trop grand (a) ou trop petit (b) entraîne un grand nombre de classes bruitées. De même un rayon d'extraction différent va impacter la qualité des classes 2D (b). Dans notre cas, le paramètre optimal se trouve autour d'un rayon d'entraînement de 54 Å et une distance minimale d'extraction de 30 pixels (c). Un plus grand nombre d'epoch augmente le score de précision, mais la valeur de ce score n'est pas liée à la qualité des classes 2D finale (a et c). Le "topaz extract" correspond à l'étapes de picking par topaz. Le rayon d'extraction est la dimension de la région dédiée à une extraction.



**Figure 2.3.1.1-2** : Les volumes *ab-initio* issus de l'approche topaz. Le chemin 1 montre les volumes obtenus à partir de l'ensemble des particules, avec en rose le volume défini comme "poubelle" et les deux autres qui montrent HTP et une densité supplémentaire en face du pertuzumab. Le chemin 2 montre les volumes issus d'une sélection de particules représentant le "doublet" de HTP.

Ensuite l'ensemble des particules sont triées via une classification 3D hétérogène (heterogeneous refinement) avec une boîte réduite à 144 voxels (Nyquist 5 Å). Les trois volumes suivants sont utilisés : Doublet de HTP, HTP, et le volume poubelle. Les 48347 particules classées dans le doublet sont fusionnées avec celles obtenues par template picking et soumises à un retrait des duplicates de particules (remove duplicate) avec une séparation minimale de 50 Å qui a retiré 40627 particules. Le topaz a donc finalement faiblement enrichi la population en doublet mais se montre performant pour trouver majoritairement les particules du doublet comme le montre la classification 2D de la figure 2.3.1.1-2. Il est aussi possible d'apprécier cette performance sur la figure 2.3.2.1.1-1.

Les particules sont triées par un "heterogeneous refinement" avec une boîte de 72 voxels (Nyquist 10 Å) dans 4 volumes définis comme suit : HTP, doublet de HTP, pertuzumab/sHER2 et un volume poubelle. Ensuite un "NU-refinement" avec les paramètres standards est lancé pour chaque classe donnant les volumes présentés dans la partie résultat (figure 2.3.2.1-1).

### 2.3.1.2 UltrAfoil

Les données de l'UltrAfoil ont été traitées sur cryosparc V3.3.2 et v4.1.2. L'ensemble des 5637 films au format tiff a été importé puis les films sont alignés via le "patch motion correction" avec les paramètres standards en utilisant toutes les images (dose totale de 60 e<sup>-</sup>/Å<sup>2</sup>). Les films sont collectés avec un taille de pixel de 0,86 Å/pixel. La CTF est estimée par le "patch CTF estimation" avec les paramètres standards. Les micrographies sont ensuite triées de la même manière que pour la quantifoil. Les coordonnées d'un "blob picking" sur 300 micrographies sont inspectées et 107642 particules sont extraites dans une boîte de 280 pixels (i.e. 240,8 Å) réduite à 70 pixels et classées dans 70 classes 2D. Les meilleurs "templates" sont utilisés pour un "template picking" sur l'ensemble des micrographies. Après inspection des coordonnées, 1014383 particules sont extraites dans une boîte de 360 pixels (i.e. 309,6 Å) réduite à 120 pixels. Après une classification 2D, plusieurs volumes *ab-initio* sont créés

séparément à partir des particules des "bonnes" et des "mauvaises" classes 2D. Les particules sont triées par trois rounds successifs de classification par "heterogeneous refinement" composé de cinq volumes dont deux sont des volumes "poubelles". Entre chaque round, la fréquence de Nyquist est augmentée en augmentant le nombre de voxels et les particules attribuées aux volumes poubelles sont retirées. 740 544 particules sont gardées pour le dernier round de classification hétérogène qui a donné les 4 volumes présentés dans la partie résultats.

### 2.3.1.3 UTC-Quantifoil

#### 2.3.1.3.1 Pre-processing et reconstruction du volume

Les données ont été traitées sur cryosparc V3.3.2 et v4.1.2. Les films au format eer ont été importés avec 54 fractions eer. L'alignement est fait en utilisant "patch motion correction" avec une réduction dans l'espace de Fourier de 1/2 et en utilisant toutes les images des films soit une dose totale de  $60 \text{ e}^-/\text{Å}^2$  et une taille de pixel de  $1,16 \text{ Å}/\text{pix}$ . La CTF est estimée avec le "patch CTF estimation" avec les paramètres par default. Les micrographies sont triées comme décrit précédemment pour en garder 8787. Les coordonnées d'un "blob picking" sur 50 micrographies sont inspectées et extraites avant une classification 2D. Huit modèles (templates) sont utilisés pour un « template picking » sur l'ensemble du jeu de données. Après inspection manuelle, 2954825 particules sont extraites dans une boîte de 280 pixels ( $313,6 \text{ Å}$ ). Après classification 2D des volumes ab-initio sont créés séparément à partir de particules issues de "bonnes" et de "mauvaises" classes. Les volumes sont ensuite utilisés pour des rounds successifs de classification hétérogène (heterogeneous classification). Après chaque round la fréquence de Nyquist est augmentée et les particules attribuées aux classes "poubelles" sont retirées. Les particules de la meilleure classe sont soumises à un affinement homogène (homogeneous refinement) ; ainsi réalignées, elles sont réextraites afin de poursuivre avec des particules centrées dans leur boîte. Ces particules sont ensuite réinjectées dans un round de classification hétérogène (heterogeneous refinement). Les particules de la bonne classe sont ensuite alignées par affinement homogène classique puis triées par classification 2D. Les 711008 particules résultantes sont réalignées par affinement homogène. La CTF des particules est corrigée globalement (Global CTF refinement) puis localement (local CTF refinement). Enfin les particules sont alignées une dernière fois par affinement homogène (homogeneous refinement), permettant l'obtention d'une carte à  $3,30 \text{ Å}$  de résolution.

#### 2.3.1.3.2 Affinements localisés

À partir de la carte haute résolution deux masques sont créés en utilisant chimeraX et "cryosparc map tool", l'un couvrant la partie pertuzumab Fab et les domaines I, II, III de sHER2 et l'autre la partie du Trastuzumab Fab et le domaine IV de sHER2. Ainsi deux affinements localisés (local refinement) sont faits à partir de ces masques en utilisant toutes les particules et sans modifier leur attribution aux deux demi-jeux de particules utilisés pour la procédure d'estimation de résolution par corrélation de Fourier.

#### 2.3.1.3.3 Reconstruction de la carte composite

Les deux cartes des affinements localisés sont fusionnées en utilisant Phenix (Afonine et al. 2018) pour créer une carte dite composite. Pour cela, il faut fournir dans l'ordre la carte de l'affinement homogène global avec les deux demi-cartes correspondantes, puis les cartes et les demi-cartes des deux affinements localisés et enfin un modèle pdb du complexe en définissant les deux régions pertuzumab/sHER2 (chaînes C, D et E résidus 1-514) et Trastuzumab/sHER2 (chaînes A, B et E résidus 515-627) correspondant aux régions des affinements localisés. La résolution est fixée à 2,5Å. En sortie l'algorithme nous donne une carte composite et les demi-cartes correspondantes. Ces cartes sont converties en format mrc par chimeraX puis importées dans cryosparc pour procéder au calcul de la FSC et de la résolution locale.

#### 2.3.1.3.4 RSPFilter

Dans le cas des données UTC-Quantifoil, nous avons converti les fichiers cryosparc du dernier affinement homogène "\*\_particles.cs" et "\*\_passthrough\_particles.cs" du format cs vers le format star avec "csparc2star.py" de la collection pyem (Asarnow et al. 2019). La commande utilisée est :

```
csparc2star.py --invert cryosparc_P22_J237_008_particles.cs  
P22_J237_passthrough_particles.cs P22_J237_particles.star
```

Ensuite nous avons utilisé RSPFilter pour générer un sous-jeu de 286000 particules (286k) à partir des informations d'alignement à savoir l'angle de rotation ("rlnAngleRot") et l'angle d'inclinaison ("rlnAngleTilt"). Ce script python permet de retirer les particules de manière aléatoire parmi les orientations les plus favorisées. Et donc de pouvoir atténuer les biais d'orientations préférentielles ou sur-représentées. Le fichier star obtenu est ensuite importé dans cryosparc, lié aux particules et utilisé pour la suite.

#### 2.3.1.3.5 Transfert vers Relion

Une fois le fichier star créé, il est possible de transférer le jeu de données vers relion pour un bayesian polishing notamment. Les grandes étapes pour lancer un bayesian polishing sur relion à partir des particules obtenues dans cryosparc sont : 1- lancer un motioncorrection sur relion, 2-convertir le fichier cs en fichier star, 3- modifier le fichier star des particules pour le faire pointer vers les nouvelles micrographies obtenues dans relion.

##### Motioncorrection dans relion

L'arborescence des dossiers de cryosparc est recrée dans l'emplacement de travail de relion. Le premier dossier contient les liens symboliques vers les films originaux. Ensuite le motioncorr est lancé sur relion.

Les films au format eer sont importés puis alignés via le "motion correction" de relion en fractionnant les films en 54 parts comprenant chacune 21 sous-images (EER fractionnement 21) correspondant à une dose de  $1\bar{e}/\text{Å}^2/\text{frame}$ .

## Préparation des fichiers pour le bayesian polishing

Les deux fichiers des données d'alignements et des chemins des particules du meilleur affinement homogène obtenus sur cryosparc sont convertis et fusionnés en format star via csparc2star comme décrit précédemment pour RSPFilter avec une commande du type :  
csparc2star.py --invert cryosparc\_P22\_J273\_particles.cs P22\_J273\_passthrough\_particles.cs fromcsparc\_particles\_swapy.star

Une fois généré, le fichier est copié et la copie est modifiée. En premier lieu le format des particules est changé de mrc vers mracs utilisant l'outil "sed" par une commande du type :  
sed -i 's/particles.mrc/particles.mracs/g' fromcsparc\_particles\_swapy\_copy.star ou en utilisant l'éditeur de texte "vi" via une commande du type :  
:%#.mrc#.mracs#

De même le nom du format des liens symboliques vers les particules est changé de mrc vers mracs par une commande du type :

```
rename 's/\.mrc$/\.mracs/' *.mrc
```

Ensuite c'est le chemin vers les nouvelles micrographies obtenues dans relion qui est modifié dans le fichier star des particules par une commande du type :

```
sed -i 's$$1/motioncorrected/$MotionCorr/job?$/RAW/$g'  
fromcsparc_particles_swapy_copy.star
```

Lorsque les films sont organisés par dossier de gridsquare, après conversion des fichiers cs en fichiers star, l'assignation des particules au dossier gridsquare correspondant est perdue. Cryosparc n'enregistre pas l'information du gridsquare alors que relion oui. Les deux fichiers star sont donc modifiés pour retirer la partie "gridsquare\_\*" du chemin menant aux micrographies. Enfin les micrographies doivent apparaître à la base de l'arborescence. Pour cela un script shell codé avec Thibault Tubiana est utilisé pour créer des liens symboliques à la base de l'arborescence. Les scripts sont les suivants :

```
#!/bin/sh  
for file in `find . -name "*.mrc"`;  
do  
  filename=`basename $file`  
  ln -s $file $filename  
done
```

Ensuite pour ramener les fichiers star :

```
#!/bin/sh  
for file in `find . -name "*.star"`;  
do  
  filename=`basename $file`
```

*In -s \$file \$filename  
done*

Ensuite les particules sont réextraites sur les micrographies dans des boîtes de 313,6 Å puis réalignées par affinement 3D. Ces particules sont ensuite traitées par "bayesian polishing" avec les paramètres vitesse 1, divergence 7000, et accélération 0,5.

#### 2.3.1.3.6 3DFlex

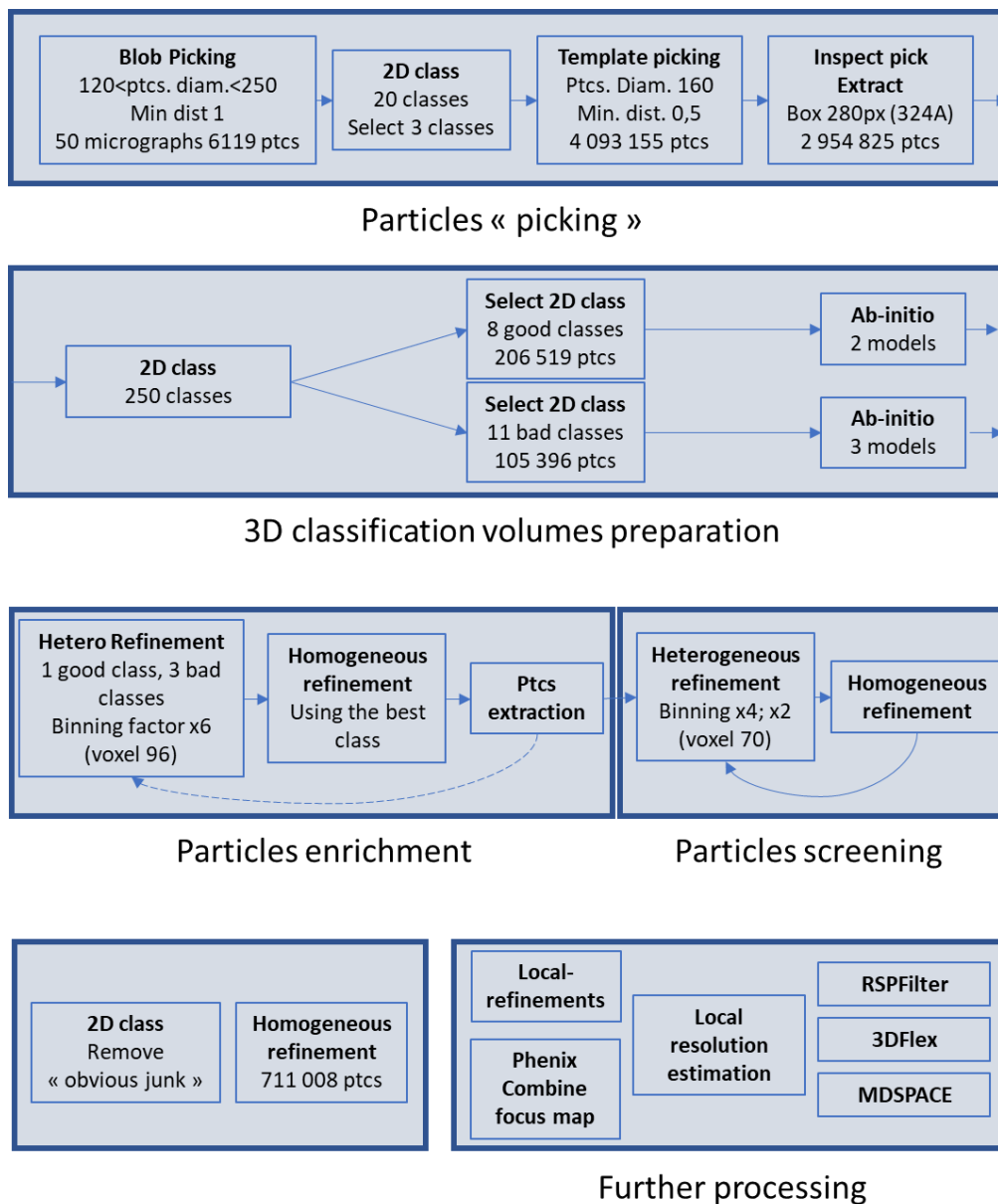
De nombreuses stratégies ont été testées avec 3DFlex. Les cartes et films présentés sont issus du traitement décrit ci-après. Nous avons exécuté l'analyse en utilisant la version 3DFlex de Cryosparc v. 4.1.2 sur l'ensemble des 711k particules et sur deux sous-ensembles de 286k particules. Un premier sous-ensemble de 286k particules provient des particules de cryosparc, le second des particules de Relion. Celles de Relion sont obtenues par extraction des 286K particules sur des micrographies alignées par l'algorithme de Relion. Les étapes du 3DFlex comprennent un recadrage de 230 à 206 pixels et un sous-échantillonnage à 100 pixels dans la préparation des données (flex data prep). Puis le masque est ajusté pendant la préparation du maillage (flex mesh prep) afin d'obtenir une maille couvrant généreusement le volume et le modèle est entraîné avec 5 dimensions latentes (flex train). La reconstruction (flex reconstruction) a été exécutée avec 60 itérations BFGS et en activant l'option d'un nouveau split des particules pour le 'gold-standard' (re-do GS split).

#### 2.3.1.3.7 MDSPACE

L'analyse par MDSPACE a été effectuée par Rémi Vuillemot et Slavica Jonic. Les particules ont été recadrées de 280 à 220 pixels et réduites pour obtenir une taille de pixel de 2,08 Å<sup>2</sup>. Deux champs de force différents ont été utilisés pour les simulations MD dans MDSPACE. Le premier est un modèle Gō à gros grain comprenant uniquement des atomes C $\alpha$ , qui a été utilisé pour analyser l'ensemble des 710k particules et étudier la variabilité globale. Le second est le champ de force CHARMM en tout atomes, qui a été utilisé pour analyser un sous-ensemble aléatoire de 5k particules et étudier la variabilité locale. Dans les deux cas, nous avons utilisé le modèle atomique dérivé de la carte obtenue avec l'affinement localisé comme conformation initiale. Le modèle a été minimisé en énergie avec chaque champ de force et les simulations ont été effectuées pendant 200 picosecondes avec un pas de temps de 2 femtosecondes en utilisant GENESIS (Kobayashi et al. 2017). Les simulations MD ont été réalisées à l'aide de l'analyse des modes normaux (NMA) par l'intermédiaire du NMMD (Vuillemot et al. 2022) pour accélérer le fitting. Les modes normaux ont été calculés à partir de la conformation initiale en utilisant un modèle de réseau élastique (ENM) comme décrit dans (Vuillemot et al. 2023) et les quatre modes normaux de plus basse fréquence ont été sélectionnés pour l'ajustement. Dans l'analyse sur le jeu de 711k particules à l'aide du modèle C $\alpha$ -Gō, MDSPACE a été exécuté pour 2 itérations, avec une première itération (entraînement) comprenant un sous-ensemble aléatoire de 1k particules, et la deuxième itération comprenant l'ensemble des données et utilisant les trois mouvements principaux appris lors

de l'itération précédente, selon le protocole itératif décrit dans l'article de Vuillemot et al. 2013. Les structures ajustées ont été alignées par rapport au pertuzumab de la conformation initiale et projetées sur un espace de faible dimension à l'aide de l'analyse en composantes principales (ACP). L'espace à faible dimension a été utilisé pour obtenir les structures moyennes présentées ainsi que les reconstructions 3D.

La deuxième analyse avec les 5k particules utilise le champ de force CHARMM, et MDSPACE a été exécuté en une seule itération. L'analyse et le calcul de la fluctuation quadratique moyenne (RMSF) se concentre sur les atomes du Fab Trastuzumab et de son interface avec sHER2 IV. L'ensemble du flux de travail MDSPACE a été exécuté avec le plugin ContinuousFlex (Harastani et al. 2022) dans Scipion (Rosa-Trevín et al. 2016).



**Figure 3.1.3-1 :** Récapitulatif et vue d'ensemble du traitement des données issues de la grille UTC-Quantifoil.



### 2.3.1.3.8 Alphafold

Les prédictions sont faites avec alphafold v2.3 sur les noeuds GPU dédiés du cluster de l'I2BC. La structure de la protéine est prédite via alphafold à partir de la séquence du domaine extracellulaire de HER2. Les prédictions d'homodimère et hétérodimère sont générées à partir des séquences suivantes :

>her2\_ecd

```
TQVCTGTMKLRPLASPETHLDMLRHLYQGCQVVQGNLELTYLPTNASLSFLQDIQEVQGYVLIAHNQV
RQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLC
YQDTILWKDIFHKNNQLALTLIDTNRSRACHPCSPMCKGSRCWGESSEDCQSLTRTVCAGGCARCKGPL
PTDCCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPY
NYLSTDVGSCTLVCPLHNQEVTAEDGTQRCEKCSKPCARVCYGLGMEHLREVRAVTSANIQEFAGCKKIF
GSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYLYISAWPDSLPLDSVFNQLQVIRGRILHNGAYSLT
LQGLGISWLGLRSLRELGSGLALIHNTLHCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGGLACHQ
LCARGHCWGPPTQCVNCSQFLRGQECVEECRVLQGLPREYVNARHCLPCHPECQPQNGSVTCFGPEA
DQCVACAHYKDPFVCVARCPGSKPDLSPYMPKFPDEEGACQPCPINCTHSCVDLDDKGCPEQRASPLT
```

>her3\_ecd

```
SEVNSQAVCPGTLNGLSVTGAENQYQTYKLYERCEVVMGNLEIVLTGHNADLSFLQWIREVTGYVL
VAMNEFSTLPLNLRVVRGTQVYDYGKFAIFVMLNYNTNSSHALRQLRLTQLTEILSGGVYIEKNDKLCHM
DTIDWRDIVRDRDAEIVVKDNGRSCPPCHEVCKGRCWGPGEEDCQTLTKICAPQCNGHCFGNPNQC
CHDECAGGCSGPQDTCFACRHFNDGACVPRCPQLVYNKLTFLQLEPNPHTKYQYGGVCVASCPHNF
VVDQTSCVRACPPDKMEVDKNGLKMCEPCGGLCPKACEGTGSGSRFQTVDSSNIDGFVNCTKILGNLDF
LITGLNGDPWHKIPALDPEKLNFRVREITGYLNIQSWPPHMHNFVFSNLTTIGGRSLYNRGFSLLIMK
NLNVTSLGFRSLKEISAGRIYISANRQLCYHSLNWTKVLRGPTTEERLDIKHNRPRRDCVAEGKVCPLCSS
GGCWGPGPGQCLSCRNYSRGGVCVTHCNFLNGEPREFAEAEFCFSCHPECQPMEGTATCNGSGSDTC
AQCAHFRDGPCHVSSCPHGVLAGKGPYKYPDVQNECRPCHENCTQGCKGPELQDCLGQTLVLIGKTHL
```

>sp|P00533|EGFR\_HUMAN\_ECD|HER1\_ECD

```
LEEKKVCQGTSNKLTQLGTFEDHFLSLQRMFNCEVVLGNLEITYVQRNYDLSFLKTIQEVAGYVLIANTV
ERIPLENLQIRGNMYYENSALAVLSNYDANKTGLKELPMRNLQEILHGAVRFSNNPALCNVESIQWRDI
VSSDFLSNMSMDFQNHGSCQKCDPSCPNWSCWGAGEENCQKLTIIICAQQCSGRCRGKSPSDCCHN
QCAAGCTGPRESCLVCRKFRDEATCKDTCPLMLYNPTTYQMDVNPEGKYSFGATCVKKCPRNYVVT
DHGSCVRACGADSYEMEEDGVRKCKKCEGPCRKVCNGIGIGEFKDSLSINATNIKHFKNCTSISGDLHILP
VAFRGDSFTHTPPLDPQELDILKTVKEITGFLLIQAWPENRTDLHAFENLEIIRGRTKQHGGQFSLAVVSLNIT
SLGLRSLKEISDGDVVISGNKLCYANTINWKKLFGTSGQKTKIISNRGENSCKATGQVCHALCSPEGCWG
PEPRDCVSCRNVSRGRCVDCNLLGEPREFVENSECIQCHPECLPQAMNITCTGRGPDNCIQCAHYID
GPHCVKTCPAGVMGENNTLVWKYADAGHVCHLCHPNCTYGCTGPGLEGCPNTPGPKIPS
```

>sp|Q15303|ERBB4\_HUMAN\_ECD

QSV CAGTENKLSLSDLEQQYRALRKY YENCEVVMGNLEITSIEHNRDLSFLRSVREVTGYVLVALNQFRYL  
PLENLRIRGKLYEDRYALAIFLNYRKDGNFGLQELGLKNLTEILNGGVYVDQNKFLCYADTIHWQDIVRN  
PWPSNLT LVSTNGSSGGRCHK SCTGRCWGPTENHCQTLTRTVCAEQCDGRCYGPYVSDCCHRECAGG  
CSGPKD TDCFACMNFNDSGACVTQCPQTFVYNPTTFQLEHNFNAKYTYGAFVCVKKCPHNFVVDSSSCV  
RACPSSKMEVEENGIKMCKPCTDICPKACDGIGTGSLMSAQTVDSSNIDKFINCTKINGNLIFLVTGIHGD  
YNAIEAIDPEKLN VFRTVREITGFLNIQSWPPNMTDFSVFSLNLTIGGRVLYSGLSLLILKQQGITS LQFQSL  
KEISAGNIYITDNSNL CYHTINWTTLFSTINQRIVIRDNRKAENCTAEGMVCNHLCSSDGCWGPDPQC  
LSCRRFSRGRICIESCNLYDGEFREFENG SICVECDPQCEKMEDGLLTCHGPGPDNCTKCSHF KDGPNCVE  
KCPDGLQGANSFIFKYADPDRECHPCHPNCTQGCNGPTSHDCIYYPWTGHSTLPQHARTP

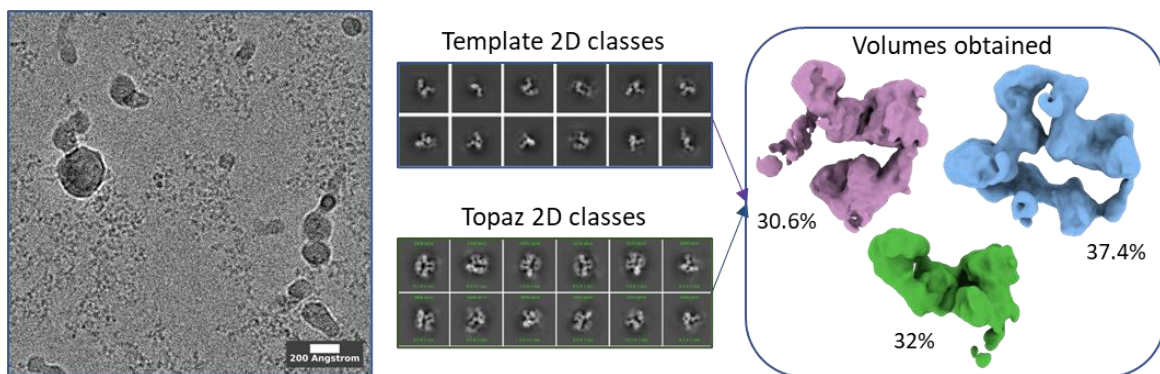
## 2.3.2 Résultats

### 2.3.2.1 Dissociation du complexe en fonction de la grille utilisée

#### Quantifoil (200kV)

Pour rappel, sur la quantifoil les particules sont très majoritairement regroupées et mal distribuées sur les 20 grilles testées. En revanche sur l'Ultrafoil les particules sont systématiquement bien distribuées sur les cinq grilles observées. Nous allons à présent nous étendre sur l'état des particules sur la grille et mettre ces résultats en vis-à-vis des résultats biophysiques.

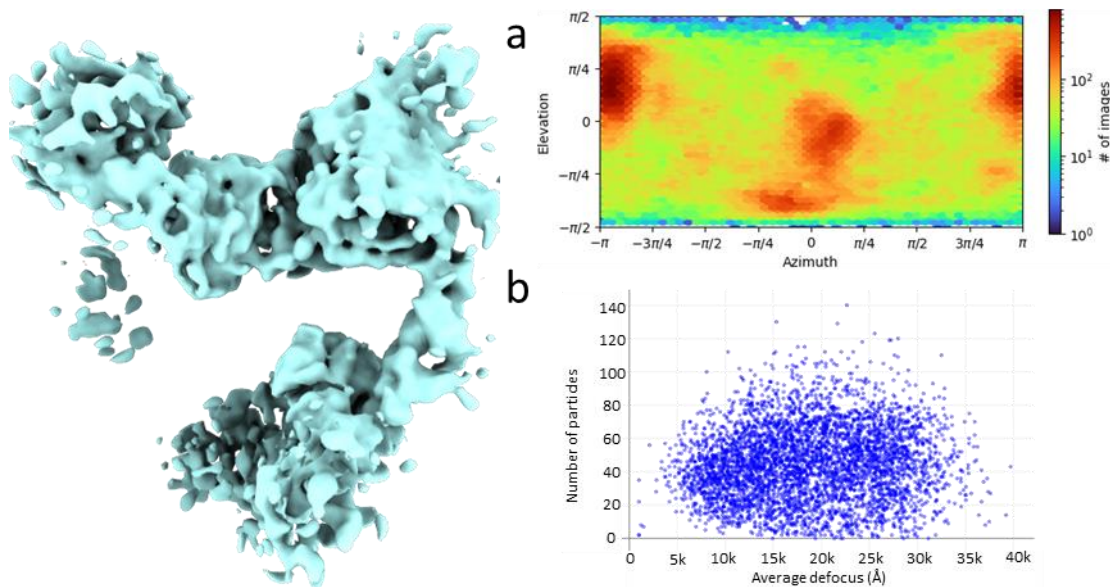
Sur la quantifoil trois volumes caractéristiques peuvent être attribués malgré la basse résolution  $\sim 7$  Å. Un premier correspondant à HTP, une forme dimérique de HTP nommée "doublet", et une forme partielle du complexe entre sHER2 et le pertuzumab nommé HP (figure 2.3.2.1-1).



**Figure 2.3.2.1-1. Quantifoil.** Bilan de l'aspect des grilles quantifoil (gauche), des classes 2D issues du protocole de template picking et de Topaz (milieu) et des volumes reconstruits avec la répartition en pourcentage des particules pour chacun des volumes reconstruits (droite). Le volume associé au complexe ternaire HTP en violet, celui au doublet HTP en bleu, et le complexe partiel entre sHER2 et le pertuzumab en vert.

La densité dans la prolongation du pertuzumab du volume HTP (figure 2.3.2.1-1 volume violet) suppose une certaine hétérogénéité au sein même de la classe. Une classification 3D sans alignement ("3D classification" de cryoSPARC) ne permet pas de mieux définir l'objet mais confirme une certaine hétérogénéité que le logiciel ne parvient pas à traiter. Enfin, bien que le volume soit reconstruit à partir de relativement peu de particules (150021 particules), des orientations préférentielles sont visibles sur le graphique des distributions angulaires sans induire d'effet anisotrope lors de la reconstruction (figure 2.3.2.1-2). De même le logiciel est capable de trouver des particules sur l'ensemble de la gamme de défocalisation (figure 2.3.2.1-2b). Les informations à moyenne résolution du doublet HTP (figure 2.3.2.1-1 volume bleu) révèlent de façon claire qu'il existe à minima une zone de contact entre les deux Fab du pertuzumab. Ce doublet pourrait être un acteur du phénomène de regroupement des particules mais certainement pas l'unique. Il est construit à partir d'un faible nombre de

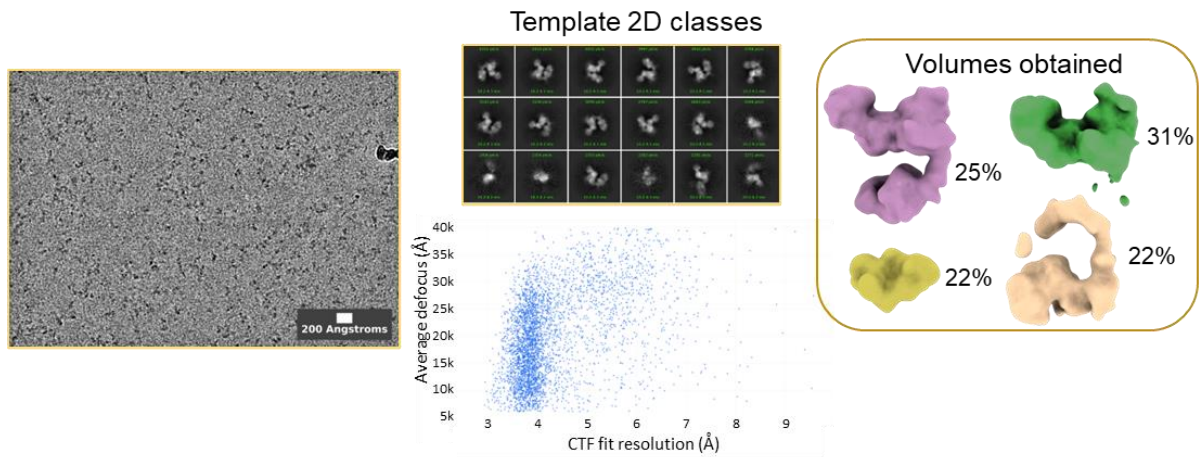
particules et souffre également d'un biais d'orientation préférentielle qui se répercute par des artefacts anisotropiques sur le volume. La zone du Trastuzumab n'est pas suffisamment définie pour être décrite avec certitude, mais il y a probablement une alternance du doublet entre des formes HTP/HTP et HTP/HP. Cette hypothèse est confortée par le volume sHER2/pertuzumab (HP) qui démontre que le Trastuzumab se décroche pendant le processus de vitrification ou au contact de la grille.



**Figure 2.3.2.1-2.** Affinement global homogène du volume HTP sur la grille quantifoil. À Gauche le volume obtenu par affinement homogène. À droite en (a) la distribution angulaire des particules. En (b) le graphique du nombre de particules en fonction de la défocalisation théorique ( $particles = f(average\ defocus)$ ).

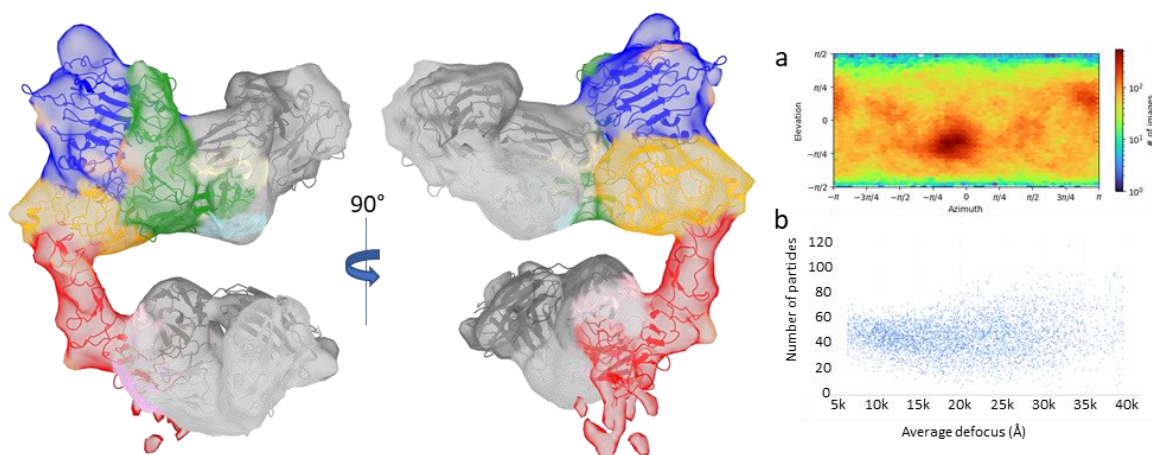
#### UltrAuroil (300kV)

La grille a d'abord eu une collecte sur un Glacios (200kV), ce qui a permis de vérifier que les particules sont bien distribuées mais a montré que l'épaisseur de la glace n'est pas optimale. Une collecte à 300kV avec filtre en énergie a donc été réalisée (cf tableau récapitulatif des jeux collectés). Dans ce deuxième jeu de données les images à faible défocalisation restent peu contrastées. Malgré cela de nouveaux éléments spécifiques à cette collecte ont pu être mis en évidence. La première observation est la présence de différents états du complexe. Nous retrouvons le volume attribuable au complexe HTP complet ainsi que celui du complexe partiel entre sHER2 et le pertuzumab déjà observé dans la quantifoil classique. Nous obtenons également deux autres volumes difficiles à identifier (figure 2.3.2.1-3).



**Figure 2.3.2.1-3.** À gauche aperçu de l'aspect des grilles UltraAufoil. Au centre, aperçu des classes 2D et du graphique de la défocalisation théorique en fonction de l'estimation de la résolution par CTF ( $\text{average defocus} = f(\text{CTF fit resolution})$ ). À droite les volumes reconstruits avec la répartition en pourcentage des particules pour chacun.

Le volume HTP est reconstruit à partir de 191677 particules. Bien que la valeur médiane de la résolution estimée par la CTF soit autour de 3,8 Å, la meilleure résolution obtenue ne dépasse pas les 6 Å, probablement limitée par l'épaisseur de la glace. Il est possible d'identifier sans peine les différentes régions de sHER2 ainsi que les deux Fabs (figure 2.3.2.1-4). La reconstruction du volume se fait avec une bonne distribution angulaire malgré qu'il y ait une petite zone d'orientations surreprésentées (figure 2.3.2.1-4a). Le nuage de points du nombre de particules en fonction de la défocalisation pour le volume HTP indique une distribution homogène des particules sur toute la gamme de défocalisation (figure 2.3.2.1-4b). En conclusion, les grilles en or permettent l'obtention de particules parfaitement distribuées mais le problème d'hétérogénéité de la population persiste.



**Figure 2.3.2.1-4.** À gauche le volume du complexe entier avec un modèle de HTP coloré par

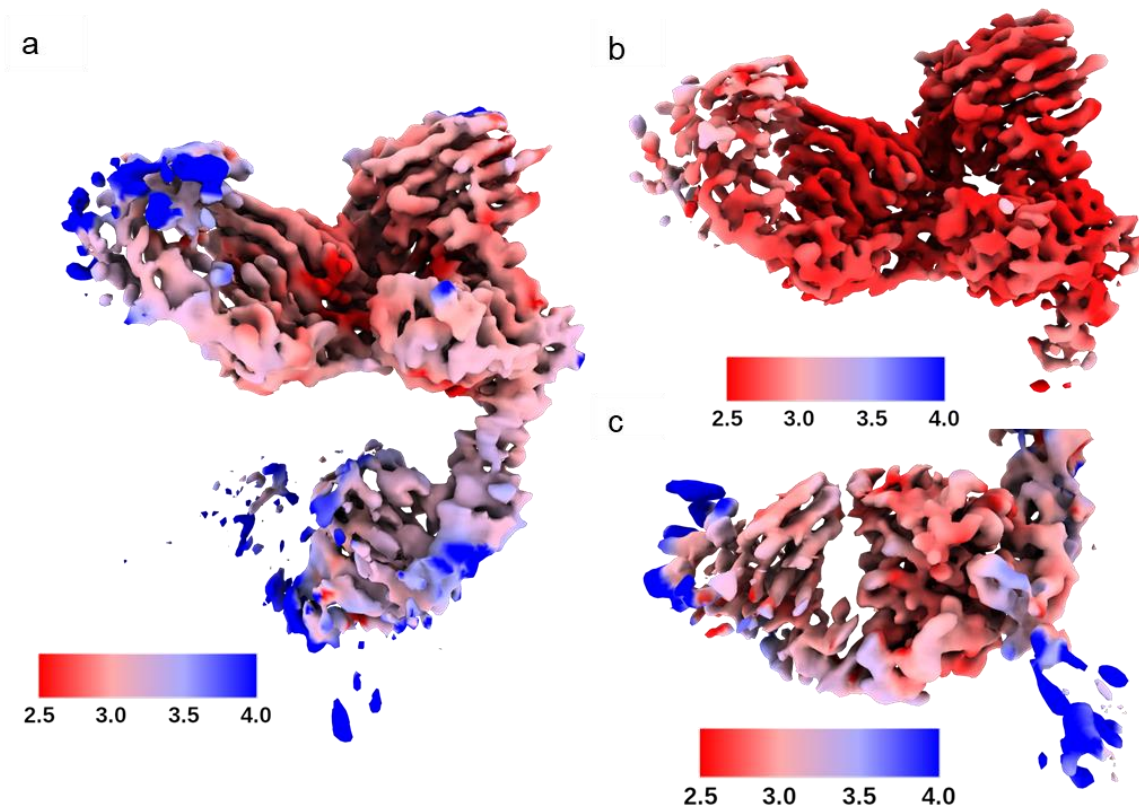
domaines (Gris foncé la chaîne lourde et gris clair la chaîne légère pour chaque Fabs, domaine sHER2 I=Bleu, II=Vert, III=Jaune et IV=Rouge) vu sous deux faces. En **(a)** la distribution angulaires des particules du volume, et en **(b)** la distribution du nombre de particules en fonction de la défocalisation.

Ces observations viennent compléter les résultats de DLS, de coloration négative et de photométrie de masse décrits dans le chapitre précédent. Notons que lors des mesures de photométrie de masse la forme doublet a été soupçonnée bien que peu intense. En revanche les formes partielles HP sont clairement visibles en photométrie de masse. De même les mesures DLS d'un échantillon ayant subi une congélation révèlent une forte hétérogénéité de l'échantillon que les analyses d'images de cryo-EM appuient. L'hypothèse d'une dégradation de l'échantillon durant le processus de vitrification est donc une piste à considérer.

### 2.3.2.2 UTC-Quantifoil, première carte haute résolution du complexe HTP

#### Une carte haute résolution

La troisième grille testée est la grille UTC-Quantifoil. Comme dit précédemment dans la partie "in-microscope" cette grille permet de manière systématique l'obtention de particules bien distribuées et une excellente densité de particules. L'analyse d'images a rapidement permis l'obtention d'une carte à une résolution de  $\sim 3,3$  Å, contre 4,4 Å pour la carte disponible jusqu'alors (Hao et al. 2019). Le protocole standard d'affinement global permet une reconstruction précise des domaines I à III de sHER2 ainsi que la région variable ( $V_H$  et  $V_L$ ) du Fab pertuzumab (figure 2.3.2.2-1a). Cependant, la carte est de bien moindre qualité dans la région comprenant sHER2 IV et le Fab du Trastuzumab, même si la carte permet déjà de visualiser une partie de l'interface. Nous avons pu améliorer la qualité de cette carte par différentes méthodes. L'amélioration la plus franche est celle obtenue par affinement localisé (local refinement), avec un premier affinement sur la zone du pertuzumab et les domaines I à III de sHER2 (figure 2.3.2.2-1b), l'autre sur la partie du Trastuzumab et sHER2 IV (figure 2.3.2.2-1c). Nous avons ainsi obtenu deux cartes haute résolution (sub-3 Å) qui ont permis de reconstruire un modèle atomique publié dans la PDB (8PHW). Ci-après, pour décrire les interfaces, nous utiliserons la même numérotation que dans la structure cristalline initiale du sHER2-Trastuzumab (Cho et al. 2003), c'est-à-dire en commençant la séquence sHER2 au résidu 1 sans tenir compte du peptide signal de 22 résidus de l'ORF HER2.



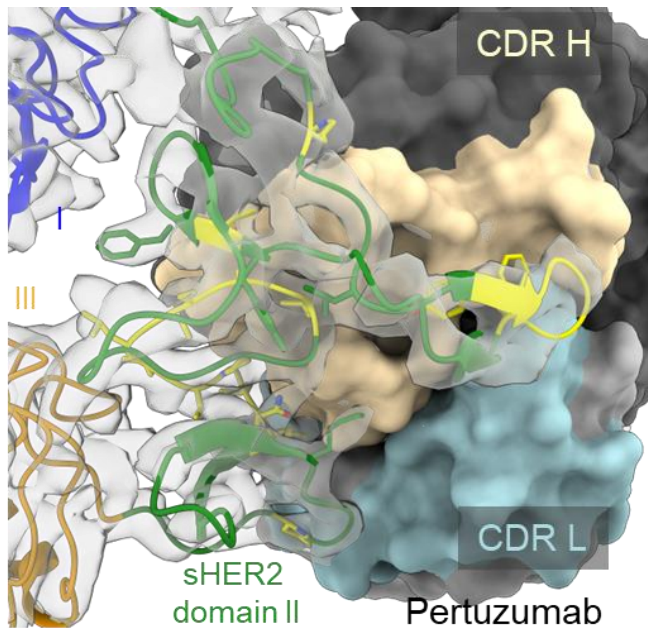
**Figure 2.3.2.2-1.** Bilan des différentes cartes obtenues. (a) carte obtenue par affinement global. (b) Carte obtenue par affinement focalisé sur la région du pertuzumab. (c) carte obtenue par affinement localisé sur le Trastuzumab.

### Analyse structurale des interfaces Fabs-sHER2

La carte d'affinement localisé permet une analyse structurale précise de l'interface her2-pertuzumab qui va être décrite en premier ci-après. Les interactions entre le pertuzumab et sHER2 sont toutes résumées et illustrées dans le tableau 3.2.2-1. Dans la région du pertuzumab, la résolution locale atteint 2,5 Å sur une grande partie de la carte (figure 2.3.2.2-1 et 2.3.2.2-2). L'interface entre HER2 et le pertuzumab est mieux définie que dans la structure cristalline disponible (1s78 (Franklin et al. 2004)). D'un point de vue structural, nous retrouvons les interactions visibles sur le modèle PDB 1s78 (Franklin et al. 2004) avec quelques variations mineures (tableau 2.3.2.2-1).

La première interaction implique le trio leucine 244, histidine 245 (sHER2) et glycine 55 (CDR H2) qui forme un empilement hydrophobe.

La deuxième région de contact implique l'interface entre le bras de dimérisation de sHER2 et



**Figure 2.3.2.2-2.** Aperçu global de l'interface pertuzumab-sHER2. Les acides aminés impliqués dans l'interaction sont colorés en jaune. La carte de l'affinement localisé est affichée en transparence. La surface du pertuzumab est colorée en cyan pour le CDR L et en jaune pâle pour le CDR H. Les domaines I, II et III sont colorés en bleu, vert et orange.

leucine 295 en face de la tyrosine 32 et la thréonine 33 (CDR H1), puis la leucine 96 et la proline 98 (CDR H3) du côté pertuzumab. Le quatrième point de contact, déjà connu, est défini par l'interface polaire impliquant notamment les liaisons hydrogènes entre la sérine 288, la thréonine 290 côté sHER2 face à l'aspartate 31 (CDR H1) et l'aspartate 53 (CDR H3). Notons qu'un mutant S288A décrit par Franklin et al. a été noté comme fortement délétère dans la liaison du pertuzumab, affaiblissant d'un facteur 6 l'interaction.

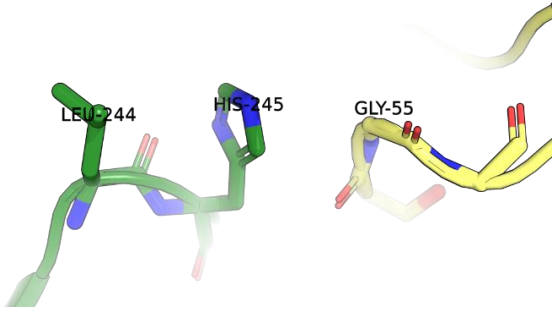
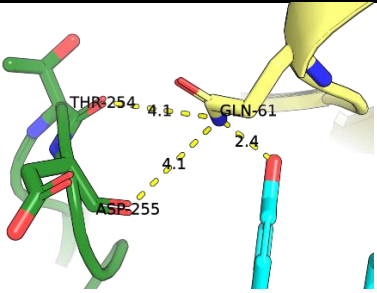
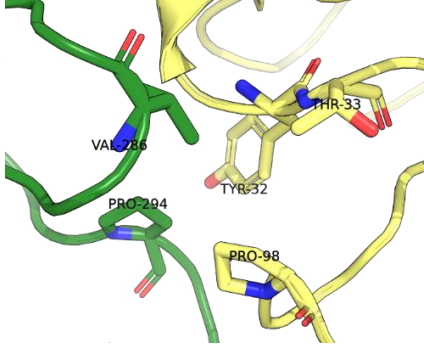
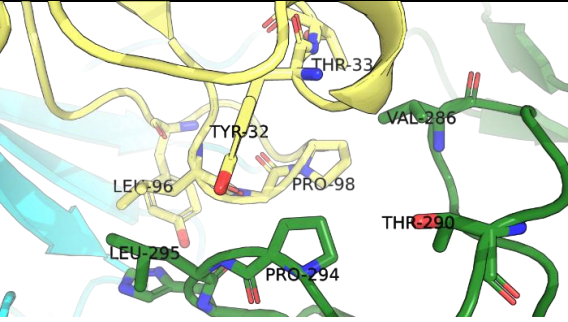
Ensuite nous trouvons une cinquième zone d'interaction définie par une liaison hydrogène entre l'histidine 296 et la tyrosine 99B (CDR H3).

C'est un empilement hydrophobe entre la proline 315 (sHER2) et les tyrosines 49 et 53 (CDR L2) qui forme la sixième et dernière interaction de l'interface entre le pertuzumab et sHER2.

le pertuzumab. Dans notre modèle nous retrouvons l'environnement hydrophobe autour de l'isoleucine 58 et la tyrosine 59 (CDR H2) ainsi que potentiellement la tyrosine 94 (CDR L3) avec la phénylalanine 257, et la tyrosine 252. De manière plus empirique nous observons également la contribution d'un cluster de résidus polaires chargés et non chargés entre les carbonyles de la thréonine 254 et de l'aspartate 255 face à la glutamine 61 (CDR H2). La lysine 64 est probablement impliqué dans une liaison hydrogène avec la glutamine 61.

Le troisième point de contact est une longue interface apolaire formée coté sHER2 par la valine 286, la thréonine 290, la proline 294 et la

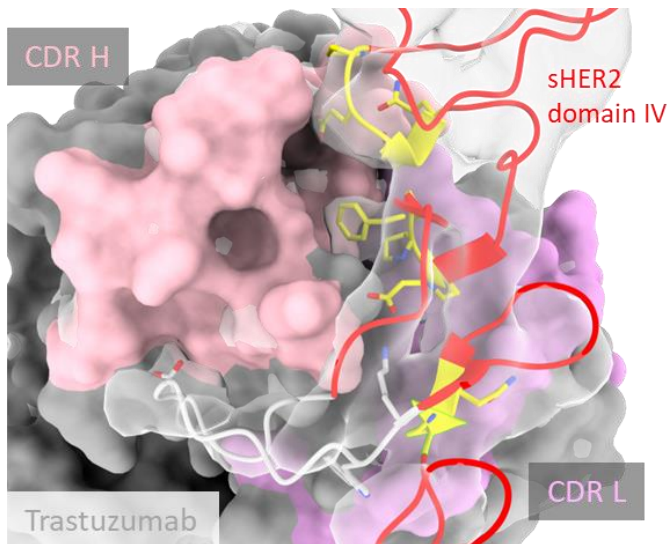


Pertuzumab-sHER2 : première zone de contact	
<p>Empilement hydrophobe : Leu244, His245 (sHER2) et Gly55 (CDR H2)</p>	
Pertuzumab-sHER2 : deuxième zone de contact	
<p>Environnement hydrophobe : Ile58 (CDR H2) Phe257, et Tyr252 (sHER2)</p>	 <p>Cluster de résidus polaires chargés et non chargés : Thr254, Asp255 (sHER2) &amp; Lys64, Gln61 (CDR H2)</p>
Pertuzumab-sHER2 : troisième zone de contact	
<p>Cluster hydrophobe : Val286 et Pro294 (sHER2) face à la Thr33, la Tyr32 (CDR H1) et la Pro98 (CDR H3)</p>	
<p>Long interface apolaire : Val286, Thr290, Pro294, Leu295 (sHER2) en face de Tyr32, Thr33 (CDR H1), Pro98, Leu96 (CDR H3)</p>	

Pertuzumab-sHER2 : quatrième zone de contact	
Interaction ionique et liaison hydrogène : Ser288, Thr290 (sHER2) face à ASP31 (CDR H1) et l'aspartate 53 (CDR H2)	
Pertuzumab-sHER2 : cinquième zone de contact	
Liaison hydrogène : His286 (sHer2) Tyr99B (CDR H3)	
Pertuzumab-sHER2 : sixième zone de contact	
Empilement hydrophobe : Pro315 (sHER2) et Tyr49, Tyr53 (CDR L2)	

**Tableau 2.3.2.2-1.** Récapitulatif des acides aminés impliqués dans les interactions à l'interface Pertuzumab-sHER2.

Bien que les deux affinements focalisés aient amélioré globalement les deux interfaces, c'est la région du Trastuzumab-sHER2 qui est la plus remarquablement améliorée (figure 2.3.2.2-1). La plupart des zones de contact décrites par Hyun-Soo Cho et al. (Cho et al. 2003) sont également identifiables sur notre carte (figure 2.3.2.2-3). Les interactions entre le trastuzumab et sHER2 sont toutes résumées et illustrées dans le tableau 3.2.2-2.



**Figure 2.3.2.2-3** : Aperçu global de l'interfaces Trastuzumab-sHER2. La carte de l'affinement localisé est affichée en transparence. Les acides aminés impliqués dans l'interaction sont colorés en jaune pour ceux préalablement identifiées (557-561, 570-573, and 602-603). Le quatrième épitope impliquant la protrusion du domaine IV de sHER2 est coloré en blanc (581-593). La surface du trastuzumab est colorée en rose pour le CDR H et en violet pour le CDR L. Le domaine IV de HER2 est coloré en rouge.

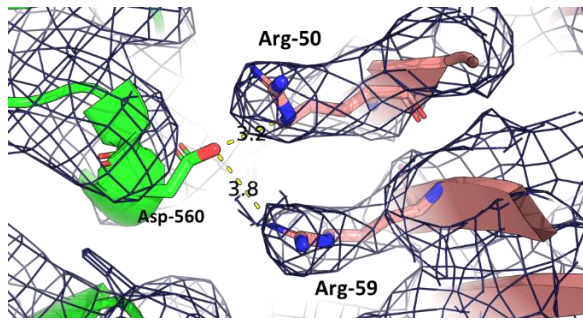
La première zone de contact débute par une interface apolaire entre la proline 557 et la tyrosine 57 (CDR H2). Dans la même zone l'arginine 59 est maintenue par deux liaisons hydrogènes avec l'aspartate 560 et la glutamine 561.

La seconde zone d'interaction se compose de deux liaisons hydrogènes entre l'asparagine 570 et l'hydroxyle de la tyrosine 105 (CDR H3) et l'amine de la lysine 593. Ainsi un cluster hydrophobe est formé coté sHER2 par la proline 572 et la phénylalanine 573 et côté Trastuzumab par le trio tyrosine 33 du CDR H1, tryptophane 99 et tyrosine 105 du CRD H3.

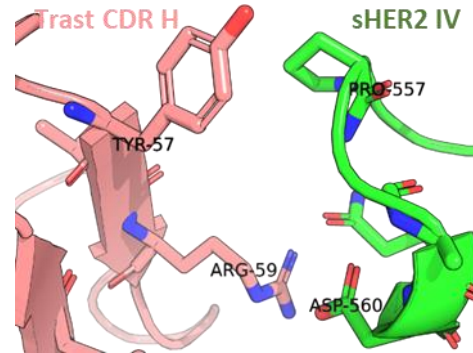
C'est au troisième point de contact où les divergences commencent. De façon empirique, le nouveau placement de la chaîne principale vient redéfinir cette zone de contact. En effet le tryptophane 592 interviendrait dans la poche hydrophobe avec la lysine 593 et la proline 603 coté sHER2 face à la phénylalanine 53, la tyrosine 49 et à la phénylalanine 104 du côté du Trastuzumab. Cependant il convient de rester prudent pour cette interprétation car aucune densité n'est visible. Cette cavité hydrophobe se poursuit du côté Fab par la leucine 46, l'alanine 34, l'alanine 106 et la tyrosine 36. C'est la proline 603 en face de la phénylalanine 53 qui vient terminer cette interface hydrophobe. Enfin nous retrouvons la liaison hydrogène entre la glutamine 602 et l'asparagine 30 du CDR L1 déjà identifiée par Hyun-Soo Cho et al. en 2003 (Cho et al. 2003).

En revanche nous proposons un nouveau quatrième point de contact à la suite de l'observation d'une densité clairement visible (figure 2.3.2.2-4). Nous pensons qu'une interaction existe entre la protrusion formée par les acides aminés 581 à 591 de sHER2 et le Trastuzumab. Nous proposons notamment que l'interaction fasse intervenir un pont salin entre l'asparagine 585 (sHER2) et l'arginine 98 du Trastuzumab.

**Trastuzumab-sHER2 : première zone de contact**

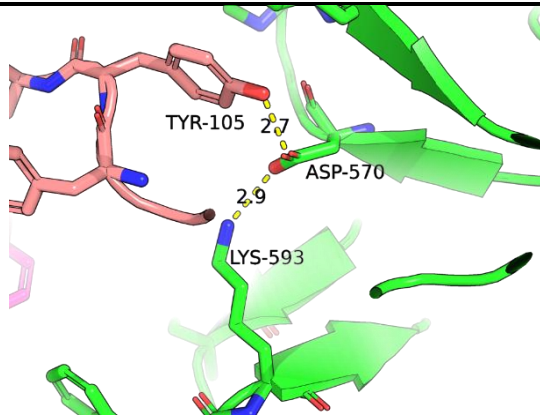


Liaison-H et interactions ioniques : Asp560, Gln561 (sHER2), Arg50 & Arg59 (CDR H2)

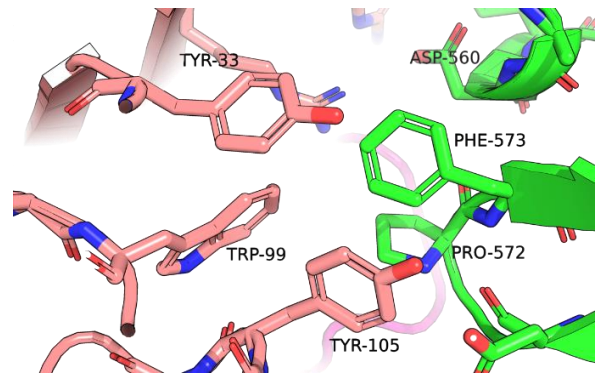


Interface apolaire : Pro557 (sHER2), Tyr57 (CDR H2)

**Trastuzumab-sHER2 : deuxième zone de contact**



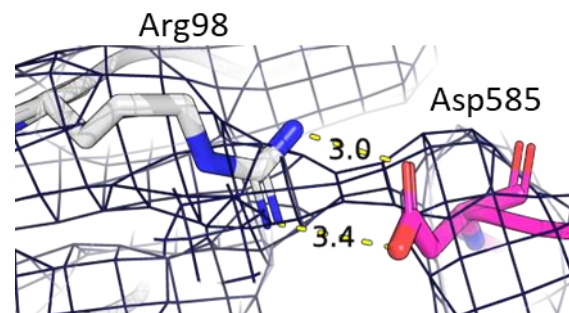
Liaisons hydrogènes : Asp570, & Tyr105 (CDR H3)



Cluster apolaire : Phe573, Pro572 & Tyr33, Trp99, Tyr105

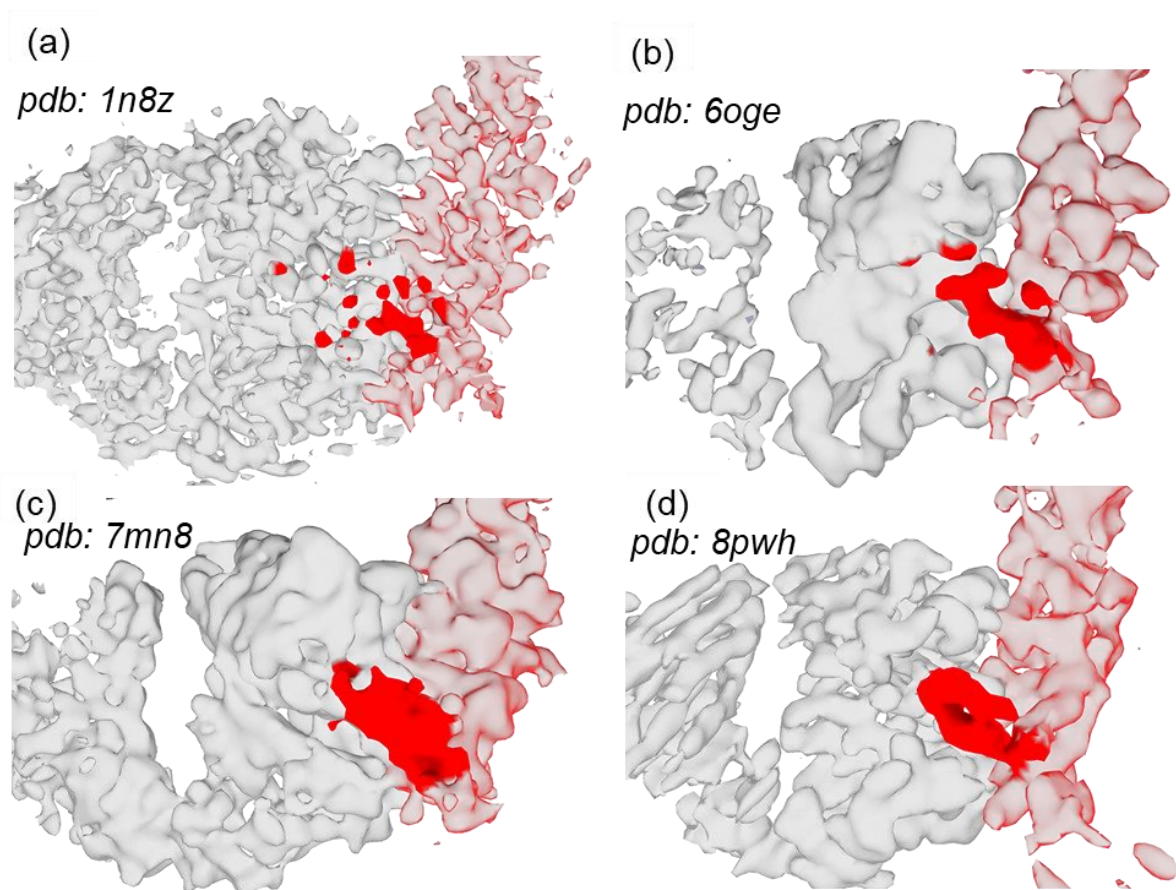
**Trastuzumab-sHER2 : troisième zone de contact**

Pont salin : Asp585 & Arg98



<b>Trastuzumab-sHER2 : quatrième zone de contact</b>	
<p>Cavité hydrophobe (vue1) : Trp592, Lys593, Pro603 &amp; Phe104 (CDR H3), Tyr49, Tyr55, Phe53, Leu46 (CDR L2)</p>	
<p>Cavité hydrophobe (vue 2) : Trp592, Lys593, Pro603 (sHER2) &amp; Phe104 Ala106 (CDR H3), Tyr49, Tyr55, Phe53, Leu46 (CDR L2)</p>	
<b>Trastuzumab-sHER2 : cinquième zone de contact</b>	
<p><b>Intéraction faible</b> : Gln602 (sHER2) et Asp30 (CDR L1)</p>	

**Tableau 3.2.2-2** : Récapitulatif des acides aminés et des interactions à l'interface Trastuzumab-sHER2



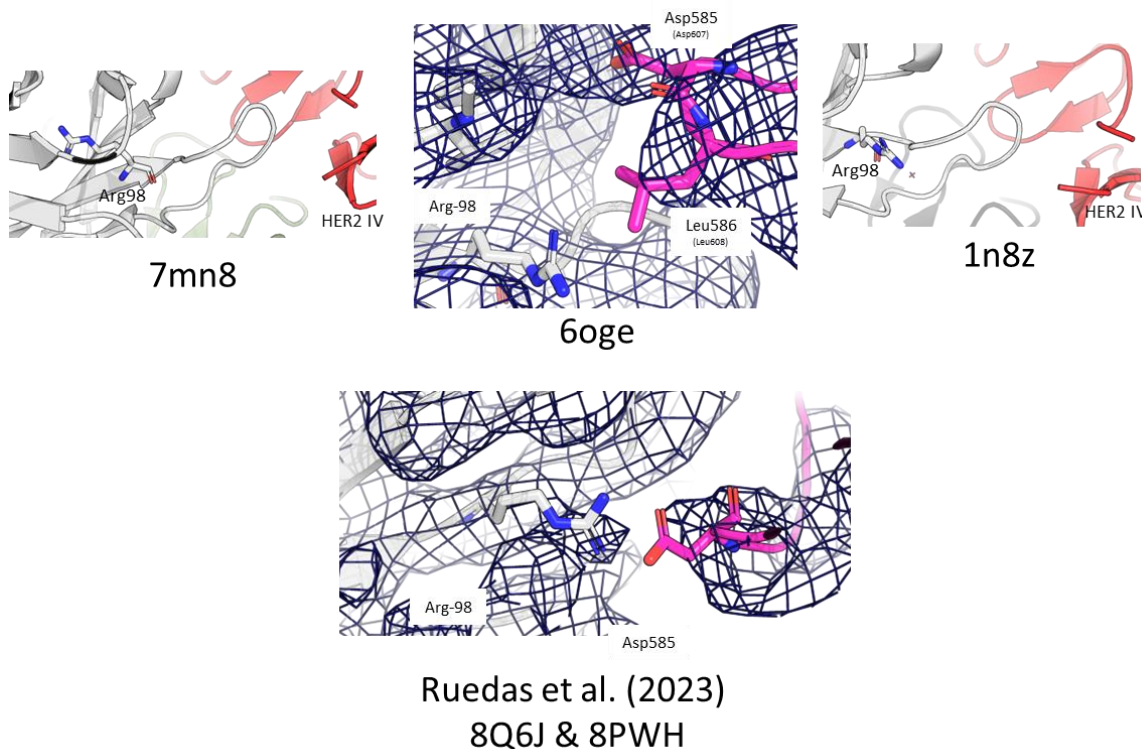
**Figure 2.3.2.2-4.** Récapitulatif et comparaison des différentes cartes disponibles dans la littérature avec la carte obtenue pendant les travaux de thèse. Les cartes **a**, **b** et **c** sont celles disponibles dans la littérature. En **d** la carte obtenue par les travaux présentés. L'interface Trastuzumab-HER2 des différentes cartes publiées. La région en rouge correspond au signal de la protrusion pour chacune des cartes du domaine IV de HER2.

La zone de la protrusion n'est pas décrite dans la littérature car soit invisible dans la carte cristallographique 2Fo-Fc (1n8z) soit faiblement visible dans les cartes cryo-EM comme celle du complexe HTP obtenues par Hao et al. (6oge) et Diwanji et al. pour le complexe HER2-HER3-NRG1 $\beta$ -trastuzumab (7mn8) (figure 2.3.2.2-4). Dans le cas de Diwanji et al. (7mn8), cette zone n'a pas été reconstruite (figure 2.3.2.2-5). Pour Hao et al (6oge), bien qu'une reconstruction de la protrusion soit proposée dans le modèle, elle n'est pas mentionnée dans l'article. La basse résolution de l'interface ne permet pas de décrire l'interface avec précision.

Tandis que pour notre carte, il y a une densité cohérente et nette dans cette zone. Certains feuillets du Trastuzumab et plusieurs départs de chaînes latérales sont clairement visibles, rendant ainsi possible la description de l'épitope et du paratope avec précision. De même, le fait de retrouver le signal de cette région sur les données expérimentales indépendantes présenté en figure 2.3.2.2-4, soutient l'hypothèse de l'existence de ce point de contact.

La position de l'arginine 98 du trastuzumab pointe vers le solvant dans tous les modèles sauf

pour une raison inconnue pour 7mn8 où l'arginine est positionnée au cœur du domaine variable V<sub>H</sub>. Dans la structure 6oge et 1n8z l'arginine 98 est déplacée pour être orientée vers la protrusion de sHER2 (figure 2.3.2.2-5). Dans 1n8z les acides-aminés de sHER2 concernés n'ont pas été construits faute de densité électronique. En revanche pour 6oge, le groupement guanidine fortement polaire de l'arginine pointe directement en direction de la leucine 608 (équivalente de la leucine 586) ce qui n'est pas idéal, et notons que la carte de donne pas de signal de la chaîne latérale. Dans notre carte, nous avons une densité claire qui permet de placer sans ambiguïté l'arginine 98 du trastuzumab, qui pointe vers la protrusion de sHER2, pour éventuellement former un pont salin avec l'aspartate 585.



**Figure 2.3.2.2-5.** Récapitulatif des différents modèles disponibles dans la PDB décrivant l'interface Trastuzumab-HER2. Le domaine IV est en rouge pour les modèles 7mn8 et 1n8z, la protrusion en rose pour 6oge et pour Ruedas et al. (2023) 8q6j & 8pwh.

<b>Récapitulatif des interfaces sHER2-Fabs</b>		
<b>sHER2</b>		<b>Fab</b>
<b>sHER2-trastuzumab interface</b>		
(1) Asp560, Gln561, Pro557		<b>CDR H2</b> Arg50, Tyr57, Arg59
(2) Asp570, Pro572, Phe573, Lys593		<b>CDR H1</b> Tyr33, & <b>CDR H3</b> Trp99, Tyr105
(3) Trp592, Pro603 Glu602		<b>CDR H3</b> Phe104, Ala106, <b>CDR L2</b> Leu46, Tyr49, Tyr55, Phe53 <b>CDR L1</b> Asp30
(4) Asp585		<b>CDR H3</b> , Arg98
<b>sHER2-pertuzumab interface</b>		
(1) Leu244, His245		<b>CDR H2</b> Gly55
(2) Tyr252, Phe257, Thr254, Asp255		<b>CDR H2</b> Lys58, Tyr59, Gln61 & <b>CDR L3</b> Tyr94
(3) Val286, Thr290, Pro294, Leu295		<b>CDR H1</b> Thr33, Tyr32 & <b>CDR H3</b> Pro98, Leu96
(4) Ser288, Thr290		<b>CDR H1</b> Asp31, <b>CDR H2</b> Asp 53
(5) His286		<b>CDR H3</b> Tyr99B
(6) Pro315		<b>CDR L2</b> Tyr49, Tyr53
<b>sHER2 domains &amp; Fab CDRs</b>		
<b>Domain I:</b> 1-185	<b>Domain III:</b> 320-485	<b>Pertuzumab :</b> CDR L1 (K24-A34); CDR L2 (S50-Y55); CDR L3 (Q89-T97); CDR H1 (G26-D35), CDR H2 (A49-G65), CDR H3 (A93-Y103)
<b>Domain II:</b> 186-219	<b>Domain IV:</b> 486-630	<b>Trastuzumab:</b> CDR L1 (R24-A34), CDR L2 (S50-Y55), CDR L3 (Q89-T97), CDR H1 (G26-H35), CDR H2 (R50-K65), CDR H3 (S97-Y109)

**Tableau 2.3.2.2-2.** Résumé des différentes interactions des interfaces entre sHER2 et les Fabs. Définition des domaines de sHER2 et des domaine CDR des deux anticorps.

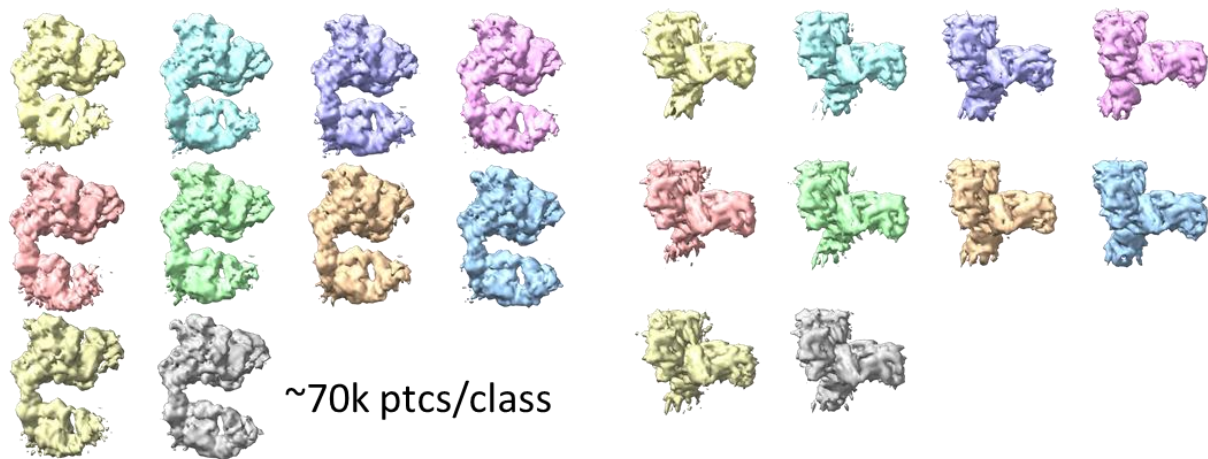
Ainsi la carte haute résolution permet de décrire simultanément et précisément les deux interfaces. Nous pouvons affirmer que la liaison concomitante du Trastuzumab et du Pertuzumab n'affecte pas l'interface sHER2-Pertuzumab, comme il avait déjà été stipulé dans l'article de (Hao et al. 2019).

### Analyse de la flexibilité par 3DFLEX et MDSPACE

Les affinements focalisés ont considérablement amélioré la résolution et ont démontré que la mauvaise qualité de la reconstruction au niveau du trastuzumab était majoritairement due à une grande flexibilité du domaine. De plus, le complexe HTP peut être organisé en deux parties. L'une comprenant le Fab du pertuzumab avec les domaines I, II et III de HER2 qui



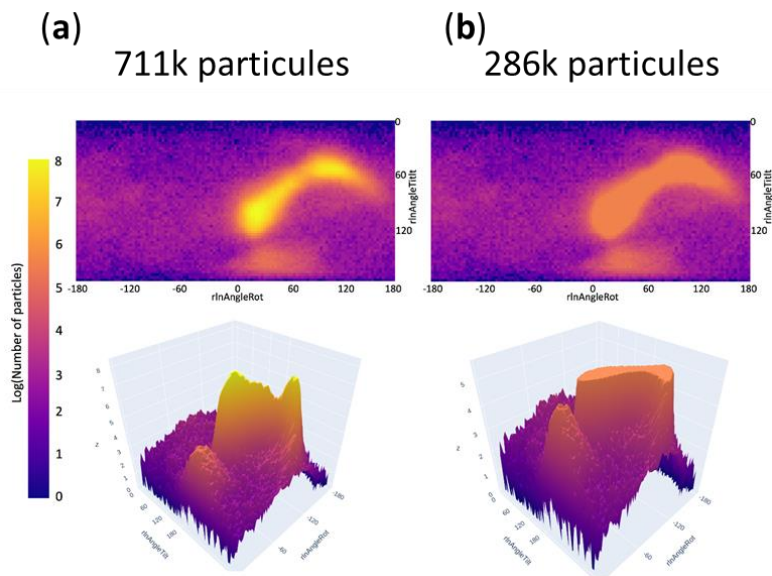
comptent pour 941 acides aminés, soit 61,5 % de l'ensemble de HTP. Tandis que la seconde partie comporte le domaine IV et le Fab du trastuzumab et contient 587 acides aminés (38,5 %). Au cours des processus d'affinement, l'alignement se fera sur la région du pertuzumab, au détriment de la région contenant le trastuzumab. Cette dernière, en plus d'être plus petite, ne peut en effet être décrite par une conformation unique. Ces affirmations peuvent être appuyées par plusieurs approches. La première est la classification 3D, bien que plutôt adapté pour classer des particules aux différences discrètes. Dans le cas de HTP nous avons procédé à une classification 3D sans alignement ("3D classification" sur cryosparc) (figure 2.3.2.2-6) qui a conduit à l'obtention de 10 classes également peuplées d'environ 71k particules présentant comme principale différence la position du trastuzumab dans l'espace, ce qui suggère une variabilité conformationnelle continue. Cette expérience montre également la forte homogénéité de l'échantillon car nous n'observons pas de complexe partiel.



**Figure 2.3.2.2-6.** Résultat de la classification 3D sans alignement. Nous obtenons 10 classes dans lesquelles sont équitablement réparties les particules. A Gauche la vue de profil et à droite la vue du dessus.

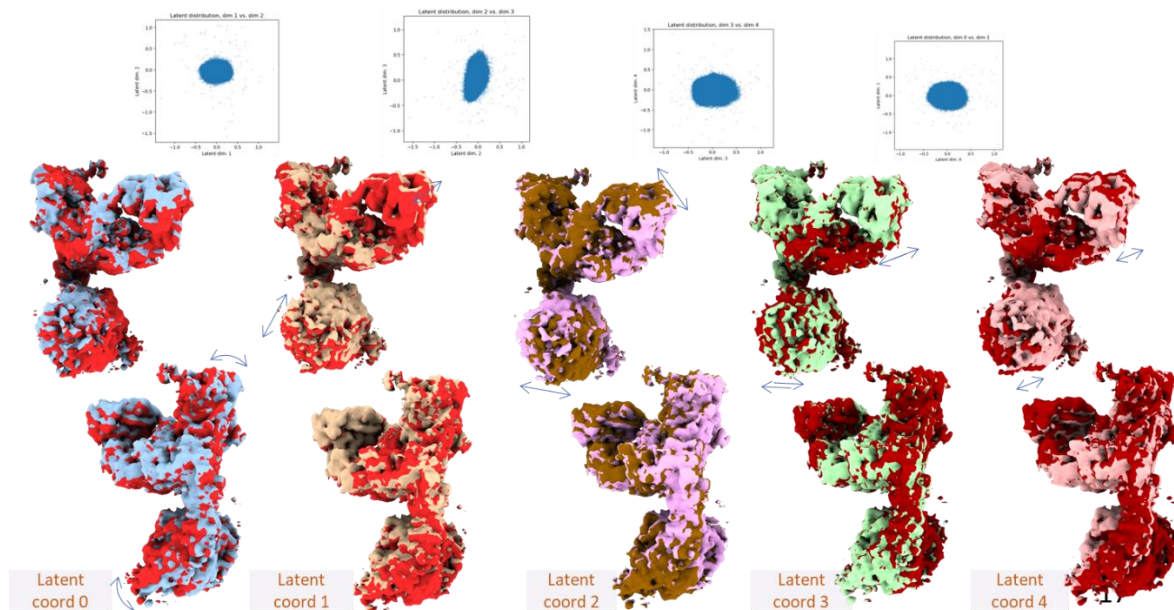
Nous avons donc appliqué des protocoles spécialisés dans l'analyse de la variabilité conformationnelle continue à partir d'images de cryo-EM. Nous avons utilisé le logiciel 3D Flexible refinement (3DFlex) (Punjani and Fleet 2021) récemment implémenté dans cryosparc. 3DFlex est un réseau neuronal de type "deep-learning" capable de modéliser l'hétérogénéité conformationnelle continue à partir des images expérimentales 2D. La particularité de cette approche réside d'une part dans sa capacité à extraire l'information de mouvement à partir des images expérimentales des particules, et d'autre part dans sa capacité à reconstruire un volume "canonique" de meilleure qualité. Ce volume est la résultante d'une convergence du signal des zones flexibles du volume des images des particules. La capacité du logiciel à rétablir les motifs haute résolution dépend du niveau d'exhaustivité du champ de déformation. Plus le mouvement décrit par le modèle est exhaustif meilleur sera le volume consensus.

3DFlex étant gourmand en mémoire vive, nous avons fait l'analyse de plusieurs manières. Une première analyse a été effectuée en utilisant l'ensemble des 711008 particules (711k) sur le cluster B3S de l'IBBC, une seconde avec un sous-ensemble de 286000 (286k) particules, sélectionnées en fonction de leur attribution angulaire avec le script RSPFilter (Figure 2.3.2.2-7).



**Figure2. 3.2.2-7.** Figure des plot RSPFilter. **(a)** 711k particules, **(b)** 286K particules. Les distributions angulaires d'un jeu à l'autre sont de plus en plus homogènes. L'échelle est la même pour **(a)** et **(b)**.

Dans l'analyse sur les 711k particules, les 5 modes de déformations montrent différents types de mouvements convaincants et attendus. Le premier mouvement est celui permis par l'angle de coude des Fabs qui entraîne un mouvement continu entre la partie constante incluant les domaines CH1 et CL et la partie variable avec les domaines VH et VL. Les modes suivants décrivent les mouvements du bloc du Fab/Trastuzumab et du domaine IV de sHER2 vers différentes directions (Figure 2.3.2.2-8).



**Figure 2.3.2.2-8** Figure montrant les différents mouvements trouvés par les différents modes de 3DFlex sur les 711k particules. Les graphiques montrent la répartition des particules en fonction de deux dimensions latentes successives. Pour chaque dimension latente nous montrons les extrêmes des volumes déformés générés par le champ de déformation de 3DFlex par deux couleurs distinctes.

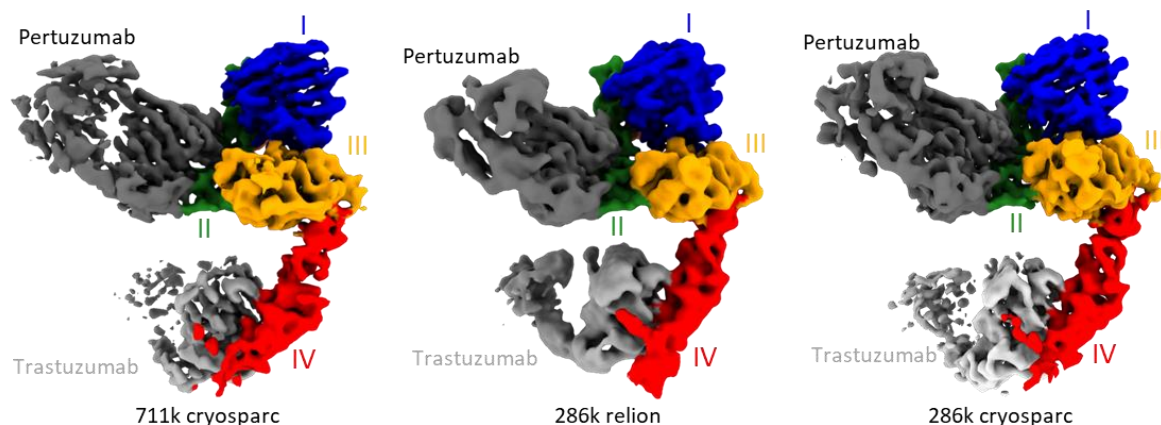
Curieusement, les mouvements décrits sont de faibles amplitudes ( $\sim 5$  Å). Nous avons donc fait varier divers paramètres dans l'optique d'améliorer l'amplitude des mouvements : une maille segmentée, des images redimensionnées à plus basse résolution, un indice de flexibilité plus permissif ou bien une initialisation des espaces latents par les résultats de 3D Variability Analysis (3DVA). Malheureusement aucun de ces éléments n'a permis d'augmenter significativement le champ de déformation et donc l'amplitude des mouvements.

Pour effectuer l'analyse sur le sous jeu de 286K particules nous avons sélectionné les particules en utilisant un script Python de Thibault Tubiana nommé RSPFilter (<https://github.com/tubiana/RSPFilter>) capable de filtrer les particules à partir d'un fichier star et ainsi retirer des particules de l'orientation la plus représentée. En travaillant sur le jeu de 286k particules, nous atténuons les biais d'orientations préférentielles et réduisons les temps de calcul ainsi que la mémoire nécessaire à 3DFlex permettant ainsi de travailler sur la station de calcul locale. Ce jeu permet de maintenir l'obtention d'une carte à une résolution convenable de 3,34 Å avec des motifs structuraux de haute résolution. (Figure 2.3.2.2-8). Avec ce sous-ensemble de particules la résolution plafonne à 4,79 Å mais la répartition des attributions angulaires est plus homogène.

Comme décrit dans le matériel et méthode, le lot de 286k particules est issu de deux provenances, une à partir d'extraction faite sur des films alignés dans relion, l'autre sur des films alignés sur cryosparc. Dans les deux cas, ce sont les mêmes coordonnées de particules.

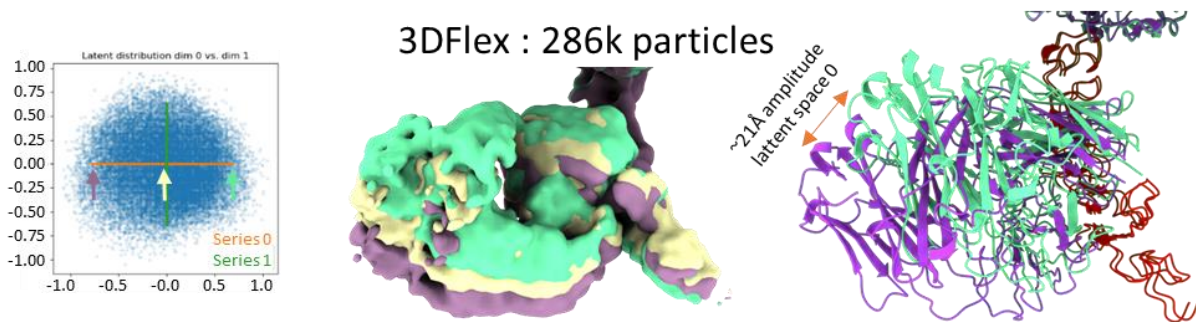
Il s'avère que ces deux lots de particules ne donnent pas les mêmes résultats sur 3DFlex malgré un protocole rigoureusement identique. Mais en réalité, le constat reste le même après différents paramétrages de 3DFlex.

Commençons par les résultats sur le sous-ensemble de 286k particules provenant des micrographies de Relion. Celui-ci montre deux choses. Premièrement l'amplitude de mouvement détectée est plus grande que précédemment. L'amplitude du balancier du trastuzumab est d'environ 21 Å (contre ~5 Å avec le jeu complet). Deuxièmement la zone du trastuzumab est mieux définie que pour le jeu complet (711k) suggérant une meilleure détection de la flexibilité par l'algorithme et donc une meilleure performance lors de la reconstruction de la carte canonique. Nous avons d'abord pensé que cette amélioration était attribuable à l'échantillonnage par RSPFilter qui avait réduit le biais d'orientation préférentielle. Mais il s'avère que ce même échantillonnage sur des particules extraites des micrographies de cryosparc ne reproduit pas les mêmes résultats. Nous avons donc effectué ici aussi une série de tests en faisant varier différents paramètres aux différentes étapes de l'analyse 3DFlex, en vain. Même si le mouvement est légèrement plus grand que celui du jeu de 711k particules, les multiples tentatives ne permettent d'obtenir mieux qu'un balancier de quelques Angströms. De même les cartes canoniques résultantes ne montrent pas d'amélioration au niveau du secteur du trastuzumab (figure 2.3.2.2-9).



**Figure 2.3.2.2-9.** Cartes canoniques obtenues par 3DFlex avec les sets de 711k particules et de 286k particules provenant de Relion et de Cryosparc.

Malgré la forte variabilité des résultats issus du programme 3DFlex de cryosparc, nous observons toujours la densité associée à la boucle du domaine IV en face du Trastuzumab. Cette observation suggère qu'aucun mouvement indépendant de la boucle du domaine IV de sHER2 n'a été détecté par 3DFlex et supporte l'hypothèse d'une nouvelle interface entre sHER2 et le trastuzumab impliquant cette boucle (figure 2.3.2.2-10). Cependant nous pouvons nous attendre à ce que le logiciel ne soit pas à même de détecter les variabilités conformationnelles d'une aussi petite zone.

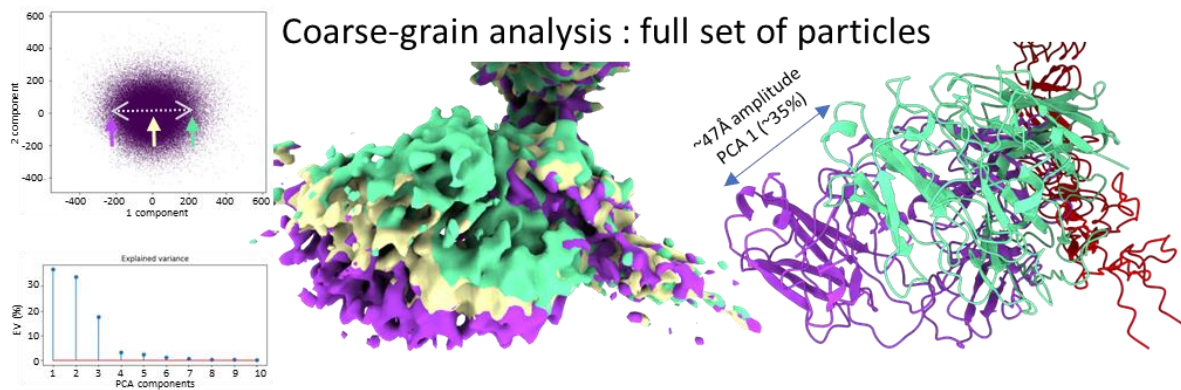


**Figure 2.3.2.2-10.** Résultat de l'analyse 3DFlex sur le 286k Relion. Le graphique à gauche montre l'attribution des particules dans deux dimensions de l'espace latent. Les lignes vertes et orange indiquent les dimensions explorées pour l'analyse. Les volumes au centre sont pris à trois moment de la dynamique, chacun correspondant à la flèche de la même couleur sur le graphique à gauche. A droite les modèles avec le domaine IV coloré en rouge et les Fabs en vert et violet en fonction de la carte dans laquelle ils ont été alignés.

3DFlex a l'avantage de récupérer l'information du mouvement de la protéine à partir des images expérimentales ainsi que de proposer un volume canonique amélioré. Mais il présente le désavantage qu'aucune connaissance préliminaire de la flexibilité sous la forme d'un modèle atomique ne peut être considérée pour l'analyse. Il présente aussi la particularité de classer les particules dans un espace latent dont il est plus difficile d'interpréter le sens physique de la dynamique 3D. L'espace latent est une notion mathématique qui permet de comparer des éléments complexes (ici des conformations) dans un graphique. C'est le bloc d'intelligence artificielle "flow generator network" qui va lier les données de l'espace latent avec l'information conformationnelle de l'objet pour chaque particule (Punjani and Fleet 2021).

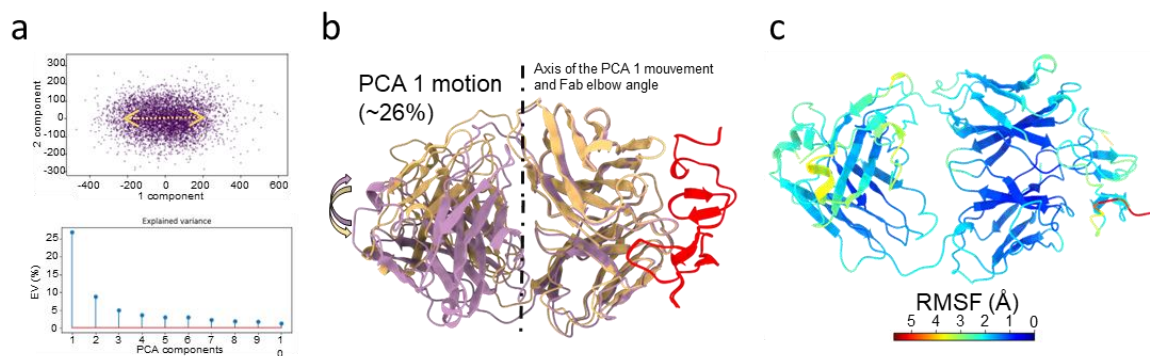
Ainsi pour confirmer la validité des résultats obtenus par 3DFlex et pour mieux caractériser la variabilité conformationnelle continue du complexe HTP, nous avons utilisé le récent logiciel MDSPACE (Vuillemot et al. 2023). Le logiciel de dynamique moléculaire et de simulation pour les analyses d'hétérogénéité conformationnelle continu en particule unique (MDSPACE) utilise un modèle atomique sur lequel les déformations obtenues par les modes normaux et les simulations moléculaires seront comparées de manière itérative aux images des particules. Par ce procédé, la simulation est liée à l'expérimental.

L'analyse faite en utilisant l'ensemble des 711k particules et un modèle gros grain montre une large amplitude de mouvements (figure 2.3.2.2-11), autour des 47 Å. Parmi ces mouvements identifiés nous retrouvons celui associé à l'angle de coude des Fabs, mais également le mouvement de balancier du trastuzumab et du domaine IV de sHER2. Les volumes ne présentent pas d'artefacts anisotropiques suggérant que les particules ont été correctement distribuées dans les différentes classes.



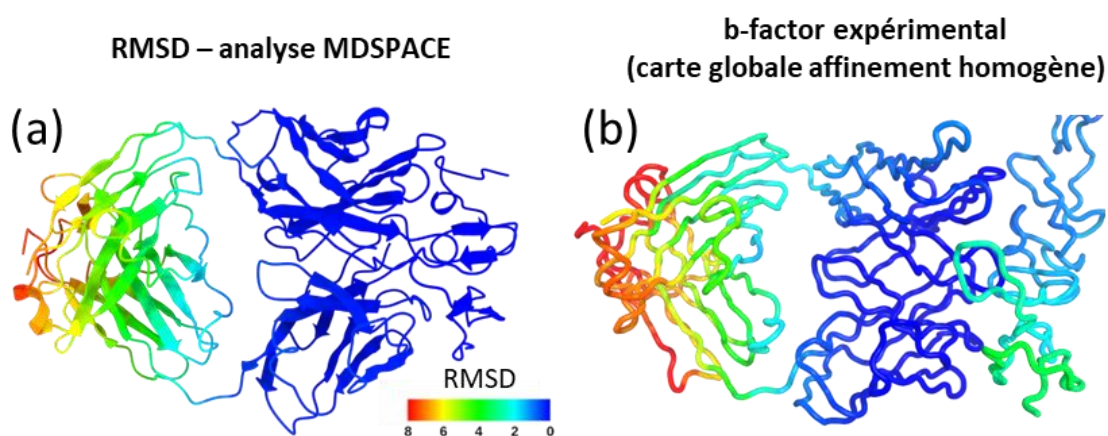
**Figure 2.3.2.2-11.** Analyse gros grain de MDSPACE sur l'ensemble des particules à pleine résolution. Les graphiques à gauche montrent l'attribution des particules selon deux composantes principales (PC1 et PC2) pour celui du haut. Le diagramme à barre en dessous montre la contribution des dix premières composantes principales. Les volumes au centre sont pris à trois moments de la dynamique, chacun correspondant à la flèche de la même couleur sur le graphique à gauche. À droite les modèles avec le domaine IV coloré en rouge et les Fabs en vert et violet associés à la dynamique PCA1 et colorés selon leur volume correspondant.

Cependant, bien que permettant une description d'un plus large espace conformationnel, le modèle gros grain ne montre aucune variation à l'interface HER2-Trastuzumab (protrusion 581-591 et CDR H3, L1 et L2). Cette observation est confortée par une seconde analyse MDSPACE focalisée sur l'interface du trastuzumab avec un modèle tout atomes et un sous jeu de 5000 particules sélectionnées de façon aléatoire (figure 2.3.2.2-12). Le mouvement principal, déductible de la première composante, et qui décrit un quart du mouvement de la zone, est celui associé à l'angle de coude entre les régions variable et constante du Fab (figure 2.3.2.2-12). En effet, lors de la simulation, les 5000 modèles (assimilable aux points de l'espace PCA) issus de MDSPACE sont alignés de manière à obtenir le RMSF montré en figure 2.3.2.2-12. Nous pouvons d'ores-et-déjà deviner une fluctuation plus importante de la partie constante du Fab par rapport au domaine variable (côté droit du Fab).



**Figure 2.3.2.2-12.** Analyse MDSPACE 5k particules. Le mouvement local de la région du trastuzumab, analysé avec un sous-ensemble de 5000 particules et en tout atome avec superposition ultérieure sur le trastuzumab. En (a), l'analyse en composantes principales : En bas, variance expliquée pour chaque composante principale. En haut, graphique bidimensionnel des deux premières composantes principales. En (b), deux modèles montrant les deux positions extrêmes de la première composante principale (26 %) après alignement sur le domaine IV de sHER2. Cela montre que le mouvement principal se situe entre les parties constante et variable du trastuzumab, conformément à l'angle de coude caractéristique des Fabs. Le panneau de droite (c) montre la fluctuation quadratique moyenne (RMSF) parmi les 5000 structures. Le C-terminus de sHER2 est la partie la plus mobile (en rouge) pour laquelle il n'y a pas de densité dans aucune de nos cartes. Nous l'avons retiré du modèle déposé à la PDB 8PWH.

Pour mieux révéler le mouvement principal associé au mouvement de l'angle de coude, nous prenons les deux modèles associés aux deux positions extrêmes identifiées à partir de l'espace PCA 1 et nous les alignons sur le domaine IV de sHER2 avant de calculer le RMSD entre les deux (figure 2.3.2.2-13a). Notons que cette observation issue de MDSPACE corrèle avec le b-facteur expérimental obtenu par l'affinement dans la carte consensus globale obtenue par affinement homogène (figure 2.3.2.2-13b).



**Figure 2.3.2.2-13.** Comparaison du RMSD de MDSPACE (a) et du b-factor expérimental (b) après affinement sur la carte issue d'un affinement homogène global.

### *2.3.2.3 Le rôle de la protrusion dans la dimérisation de sHER2 par des analyses alphafold. Les prédictions alphafold des multiples partenaires de sHER2*

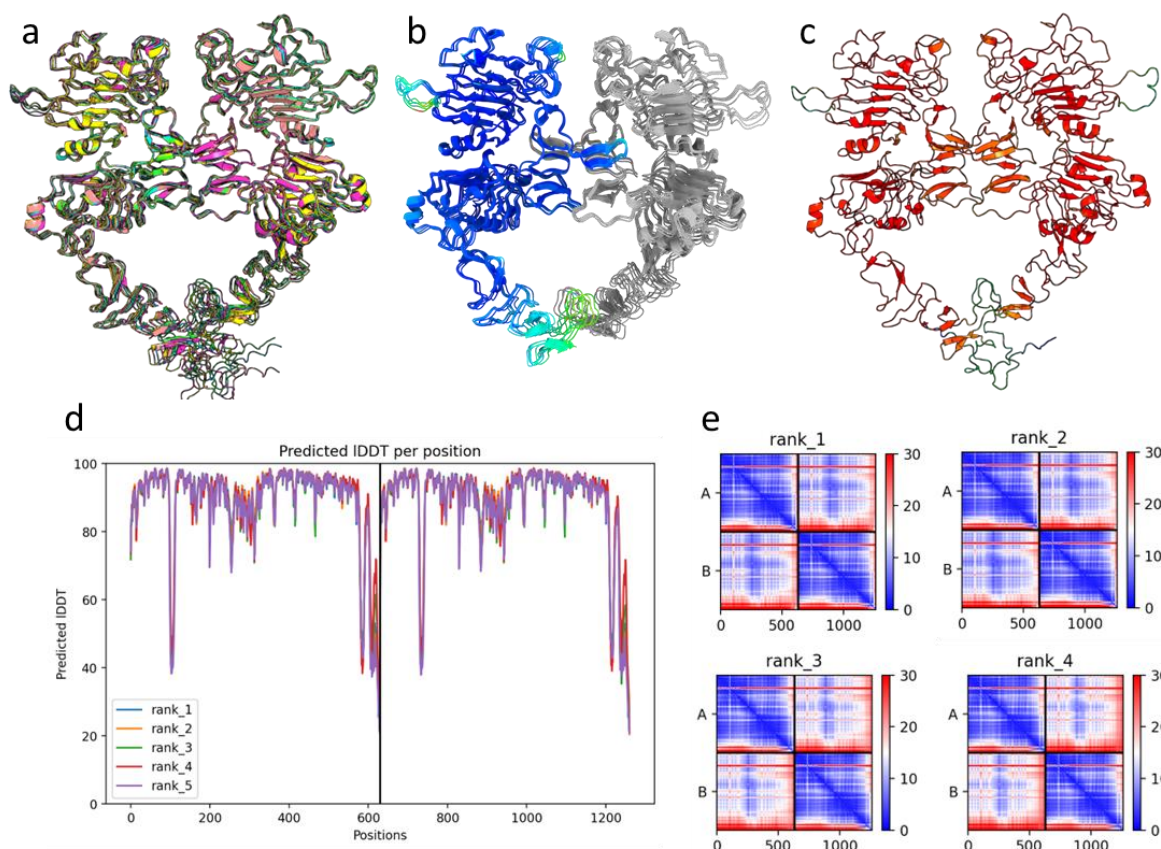
La protrusion du domaine IV de sHER2 est structuralement proche de celle présente dans le domaine II mais n'est pas connue pour être impliquée dans un processus de dimérisation. Or, un tel rôle impliquera une inhibition de ce point de dimérisation par le trastuzumab et pourrait donc expliquer la synergie lors de son utilisation avec le pertuzumab. Nous allons donc au travers de ce chapitre investiguer cette hypothèse avec l'outil alphafold.

Bien qu'historiquement non conçu pour cette fonction, Alphafold se révèle aujourd'hui être un bon outil pour aider à argumenter sur la véracité des oligomérisations. Nous allons illustrer une des façons dont cet outil s'insère dans le paysage des études structurales. L'une des particularités de la famille des ErbB est la possibilité de former de multiples combinaisons de dimérisation entre ses 4 membres (figure 1.2.1-1). L'argumentaire de l'existence de ces dimères à longtermes reposé sur des données de biologie cellulaire et est très peu documenté par des données structurales. Ce n'est que récemment en avec la publication des cartes cryo-EM 7mn6-8 (Diwanji et al. 2021) et 8hgs (Bai et al. 2023) que nous disposons de données structurales qui révèlent les mécanismes de dimérisation. Cependant il reste encore un doute quant à l'intervention de la protrusion 581-591 dans la dimérisation.

Deux méthodes sont utilisées pour estimer la fiabilité d'un modèle prédit par alphafold. La première est test de différence de distance locale (LDDT), c'est un score qui évalue les différences de distance entre les atomes d'un modèle et une référence. Elle est bien adaptée à l'évaluation locale de la qualité des modèles (Mariani et al. 2013). Le deuxième est la prédiction d'erreurs d'alignement (PAE), qui indique l'erreur de la position entre deux résidus. Il est calculé pour l'ensemble des paires de résidu possible (Varadi et al. 2021).

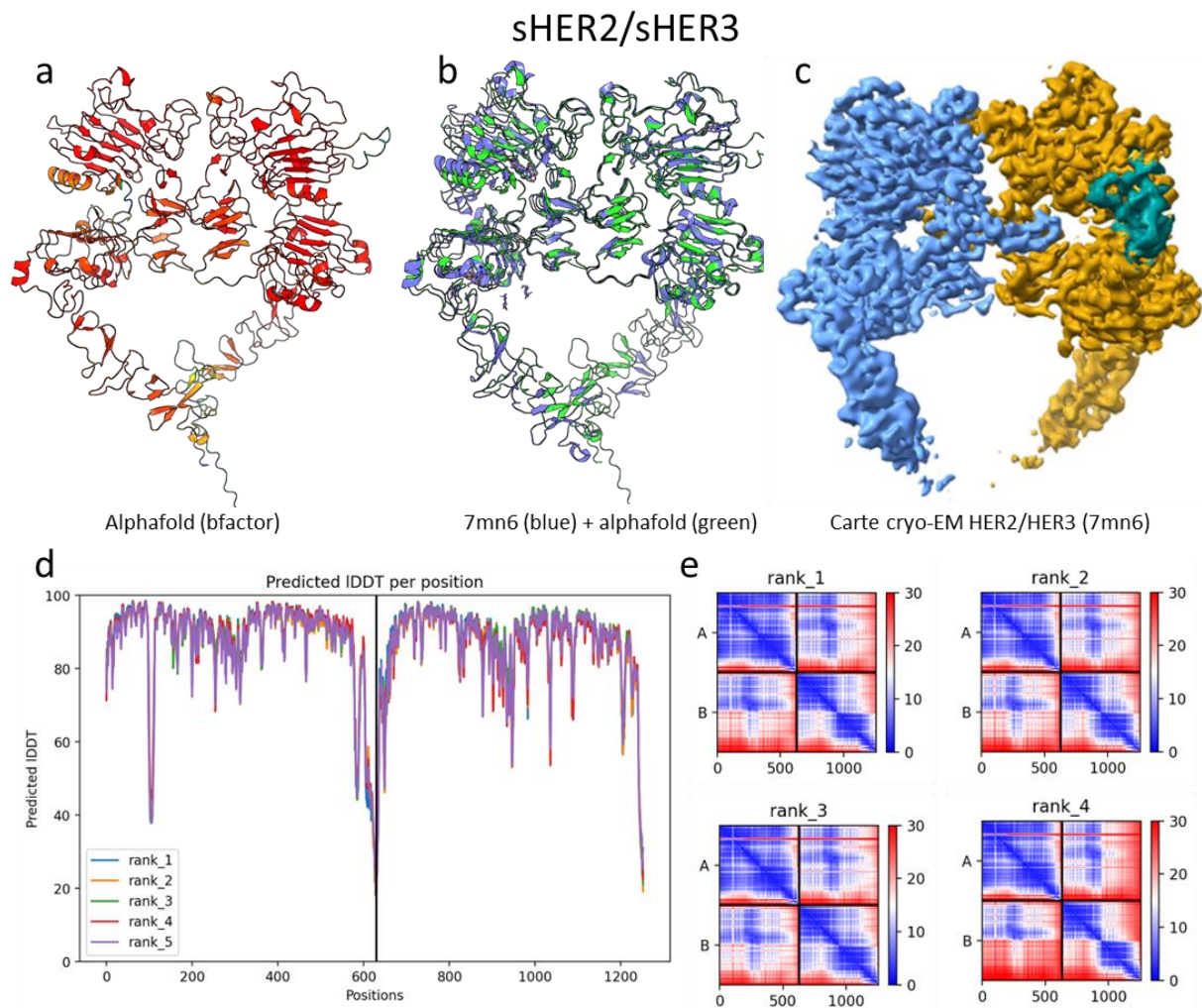
Les prédictions alphafold de l'homodimère sHER2-sHER2 présentent un score de fiabilité très correct. Excepté niveau de la protrusion où les scores LDDT et PAE sont les plus faibles (figure 2.3.2.3-1). En revanche le score global PAE sur les autres domaines est très correct et semble même meilleur que pour l'hétérodimère sHER2/sHER3 (comparer la figure 3.2.3-1e avec la figure 2.3.2.3-2e).





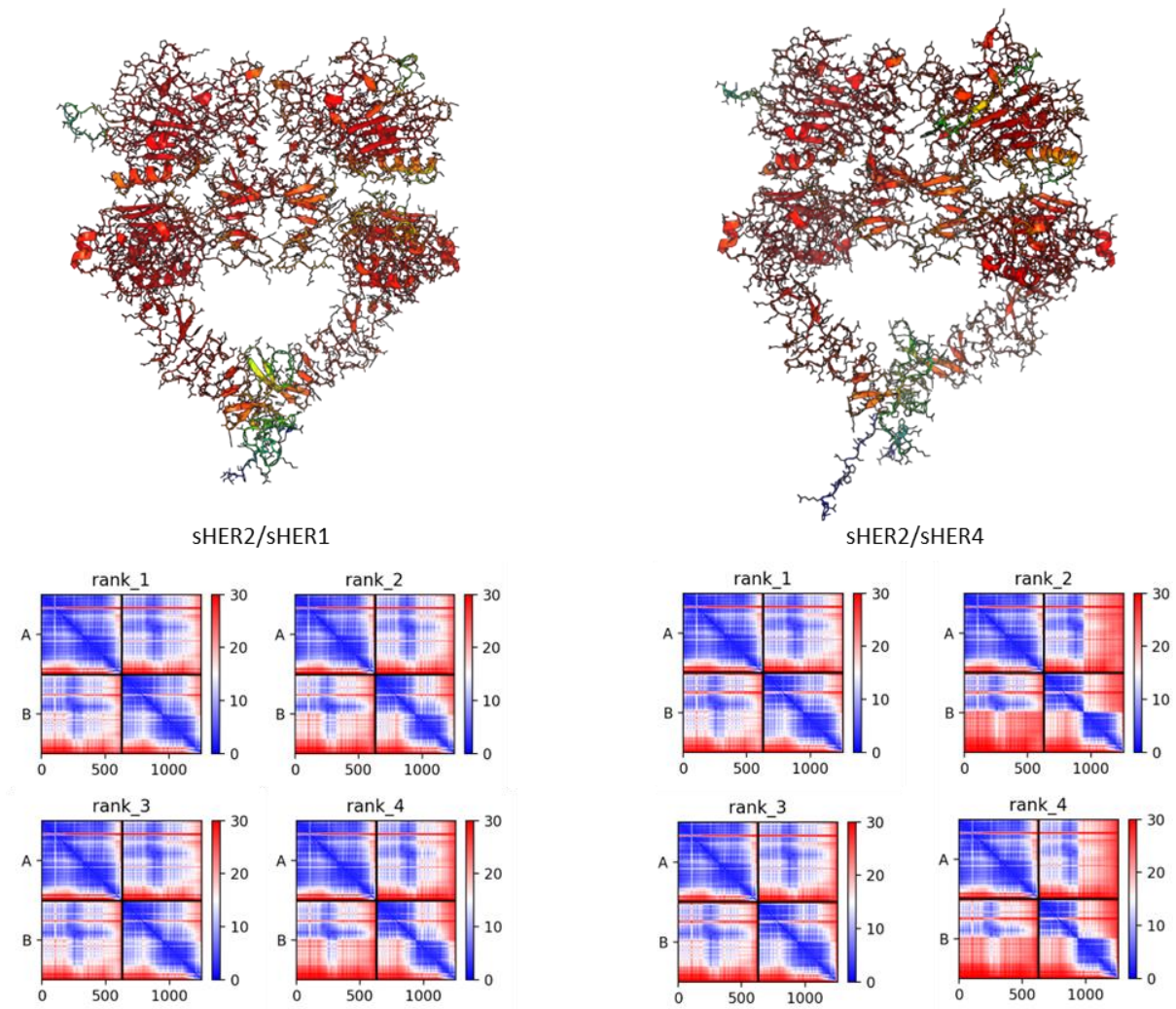
**Figure 2.3.2.3-1.** Modélisation alphafold de l'homodimère sHER2. A gauche (a) les cinq prédictions superposées. Au milieu (b) même image colorée par RMSD le c-ter présente une forte variabilité. A droite (c) la prédiction ayant le meilleur score coloré par score de prédiction du plus haut indice de confiance au plus faible du rouge au bleu respectivement. En bas à gauche (d) le graphique des scores LDDT des différentes prédictions et à droite (e) les scores PAE des quatre premiers modèles.

Les modélisations d'hétérodimères HER2/HER3 ne montrent pas de signe de dimérisation par le domaine IV. En effet le domaine IV présente une forte variabilité d'un modèle à l'autre et les score PAE sont encore inférieurs à ceux obtenus pour l'homodimère (figure 2.3.2.3-2). Les modélisations avec le domaine IV uniquement ne sont pas convaincantes, les domaines s'agencent aléatoirement l'un par rapport à l'autre (non montré). Notons que sHER3 est soupçonné d'avoir deux conformations (active et inactive) selon qu'il soit ou non activé par un EGF (Wieduwilt and Moasser 2008). AlphaFold ne prédit qu'un repliement semblable à la forme active capable d'interagir avec un partenaire. Cette modélisation peut être appuyée par les données expérimentales de la structure cryo-EM de l'hétérodimère HER2/HER3/NGR1 $\beta$  (pdb : 7mn6 (Diwanji et al. 2021)). La particularité de cette structure réside dans l'utilisation des deux protéines avec leur partie transmembranaire incorporée dans une micelle de détergent Dodecyl-D-Maltoside (DDM). Ceci permet de maintenir les deux parties C-terminales rapprochées, maintenant les protrusions du domaine IV face à face. Malgré cela, aucune densité claire n'est visible à ce niveau du domaine IV (figure 2.3.2.3-2c).



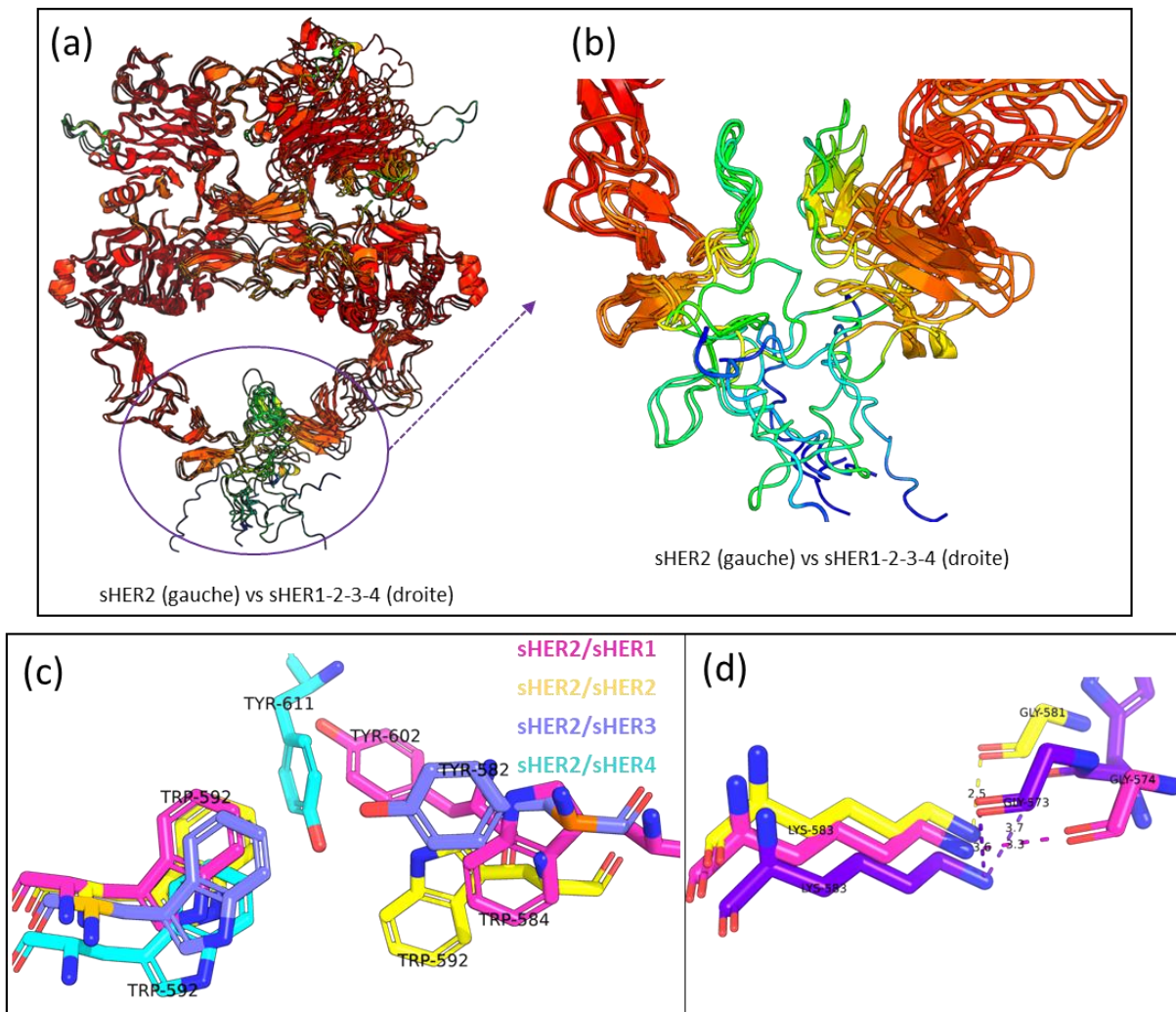
**Figure 2.3.2.3-2.** Prédiction alphafold et résultats expérimentaux de l'hétérodimère HER2/HER3. A droite (**a**) le modèle prédit colorié par score de fiabilité. Au centre (**b**) le modèle alphafod en bleu superposé sur la structure expérimentale en verte. A droite (**c**) la carte cryo-EM du dimère sHER2-sHER3/NGR1B. En **d** et **e** les graphiques du score LDDT par résidu et les scores PAE des quatre premiers modèles.

Enfin concernant les hétérodimères sHER2/ERBB et sHER2/sHER4 ils donnent les mêmes résultats. De manière systématique les domaines I à III ont de très bon scores LDDT et de PAE. En revanche le PAE est meilleur pour sHER2/sHER1 que pour sHER2/sHER4. Il est très similaire à celui obtenu pour la prédiction sHER2/sHER3 qui est appuyée par des résultats expérimentaux. Le domaine IV obtient des scores plus faibles aux niveaux de la protrusion 581-591 et de la partie c-terminale.



**Figure 2.3.2.3-3.** Prédiction alphafold pour les dimères sHER2/ERBB (sHER1) et sHER2/sHER4. En haut les modèles de rang 2 avec sHER2 à gauche et sHER1 ou sHER4 à droite.

Les modèles obtenus avec alphafold ne permettent pas d'appuyer sérieusement l'hypothèse d'une éventuelle implication de la protrusion dans les processus de dimérisation. Bien que celle-ci soit souvent placée à proximité de son partenaire il est difficile de conclure quant à son effective implication. Il est possible que cette proximité soit une résultante d'une relative rigidité qui induit un vis-à-vis de cette zone. La seule relative constance dans la prédiction de l'interface entre sHER2 et ses partenaires est la lysine 583 de sHER2. Celle-ci pointe vers le carbonyle d'une glycine, à savoir la glycine 574 pour sHER2/sHER1, la glycine 581 pour sHER2/sHER2, et la glycine 573 pour sHER2/sHER3. Mais cette observation n'est valable que pour certaines des prédictions, généralement les prédictions de rang 2. Pour les autres rangs cette zone n'est pour la plupart pas interprétable à cause de clash. De même le tryptophane 592 qui dans notre modèle pourrait être impliqué dans une interface avec le Trastuzumab, est dans certains modèles placés en face d'une tyrosine bien que les positions et les distances ne soient pas idéales pour une interaction apolaire (figure 2.3.2.3-4). Enfin il n'y a pas de signal d'interaction de l'asparagine 585 ou de la Leucine 586 qui sont des acides aminés clés dans l'interaction avec le trastuzumab.



**Figure 2.3.2.3-4.** Superposition des prédictions alphafold des différents partenaires possibles entre sHER2 et les autres récepteurs EGFR. En haut en (a) la vue globale des prédictions alphafold entre sHER2 et ses différents partenaires, alignées sur sHER2. Cette représentation permet d'apprécier la constance de l'interaction impliquant les domaines I et II. En (b) un zoom sur la protrusion 581-591 dans laquelle on observe la fluctuation spatiale du placement du partenaire de sHER2 au voisinage de la protrusion 581-591. En dessous en (c) la proximité du tryptophane 592 de sHER2 avec les aromatiques du partenaire et en (d) un zoom sur la lysine 583. Les différentes prédictions sont colorées différemment. À droite l'interface impliquant les lysines. Les codes couleur sont les mêmes.

### 2.3.3 Conclusion et discussion : nouvelle structure du complexe HTP et analyse de la flexibilité continue

Dans le chapitre précédent nous avons vu à quel point le changement de grille pouvait impacter le comportement de l'échantillon. Nous avons décrit le lien étroit entre la nature de la grille et le comportement de l'échantillon depuis la distribution des particules, au biais d'orientations préférentielles et à l'état du complexe sur la grille. Des résultats extérieurs à ma thèse, obtenus par Eduardo Bruch et Carla Cabal-Fernandez, ont montré que le complexe sHER2-Pertuzumab avec un nanobody contre sHER2 ou contre le Fab se comporte bien sur grille quantifoil. La seule différence provient donc de l'absence du trastuzumab. Même si le doublet semble indiquer un point de contact possible entre les pertuzumab, il se pourrait que ce soit en fait le trastuzumab qui soit à l'origine des difficultés rencontrées. Bien que la distribution et l'orientation des particules soient des problèmes connus en cryo-EM, il est moins commun de favoriser le changement de la nature de la grille pour y remédier.

L'obtention de notre carte haute résolution montre que la flexibilité d'un échantillon n'est plus une limitation absolue en cryo-EM. Cela reste encore une contrainte, mais nous avons montré que dans de bonne condition il est possible de retrouver le signal d'une petite zone flexible de 50 kDa. Cette carte a permis d'éclaircir les interfaces entre les deux Fabs et sHER2. Mais surtout celui entre le trastuzumab et sHER2. La protrusion définie par les acides aminés 581-591 du domaine IV présente un repliement très similaire au bras de dimérisation du domaine II. Malgré ces ressemblances, la zone n'a pas de rôle clairement identifié, bien qu'elle fût anciennement proposée comme impliquée dans l'homo-dimérisation ou l'hétéro-dimérisation des récepteurs. Compte tenu des récentes structures et des modèles alphafold, il n'est pas impossible que cette région crée une zone d'interaction faible, qui serait secondaire à la dimérisation principalement médiée par la rencontre des domaines II. Ce scénario expliquerait aussi l'effet synergique qui résulte de l'utilisation des deux anticorps.

Nous avons également montré la puissance des derniers logiciels de traitements de la flexibilité continue en cryo-EM. Chacun ayant ses forces et ses faiblesses, par ces travaux nous avons montré qu'ils sont capables de révéler les mouvements complexe de plusieurs dimensions.

La densité est difficile à améliorer du fait de la variabilité conformationnelle continue de la zone et de sa relative petite taille. Il est envisageable d'améliorer la carte cryo-EM de cette zone en faisant une collecte sur un microscope électronique de 300kV par exemple. Une autre approche pourrait être d'artificiallement faire grossir la partie trastuzumab/sHER2 en y fixant un partenaire. L'utilisation de nanobody a déjà fait ses preuves pour ce genre d'application. Plus original, l'utilisation de protéines artificielles telles que les alpharep (Urvoas et al. 2012) ou les Darppin (Stumpp et al. 2008) peuvent aussi être une option. Idéalement, le partenaire pourrait se fixer sur le domaine constant du Fab pour limiter au maximum le risque de perturber l'interaction. Nous pouvons aussi imaginer un partenaire qui se fixe sur le domaine IV mais sur la face opposée de l'interface. Les outils de caractérisation biophysique

permettront aisément de sélectionner les partenaires les plus prometteurs avant de faire des images de cryo-EM. En revanche ce ne sera plus le complexe "natif" mais une version "artéfactuelle".



### 3 La protéine de non structurale 4 (NS4) de norovirus



Caractérisation biochimique, biophysique et prédictions bio-informatiques de la protéine NS4

Des Fabs pour une structure par cryo-EM de la protéine NS4

Les Alphareps dirigées contre la protéine NS4





### 3.1 Caractérisation biochimique, biophysique et prédictions bio-informatiques de la protéine NS4

La protéine NS4 n'est pas précisément décrite dans la littérature et les informations relatives à son comportement biochimique sont inexistantes. De même il n'existe pas de donnée structurale expérimentale. Dans ce chapitre la protéine NS4 de référence est la version sauvage entière du génogroupe II génotype 4 variant Helensbourg qui provient d'un isolat du variant New Orleans pandémique à l'hiver 2009-2020. Elle sera nommée NS4<sub>FL</sub> ou NS4 dans le texte.

#### 3.1.1 Matériel et méthode : clonage, expression et purification de NS4

Le protocole d'expression et de purification de NS4<sub>FL</sub> a été mis en place par Johann Habersetzer (travaux interne non publié) et repris avec des modifications mineures.

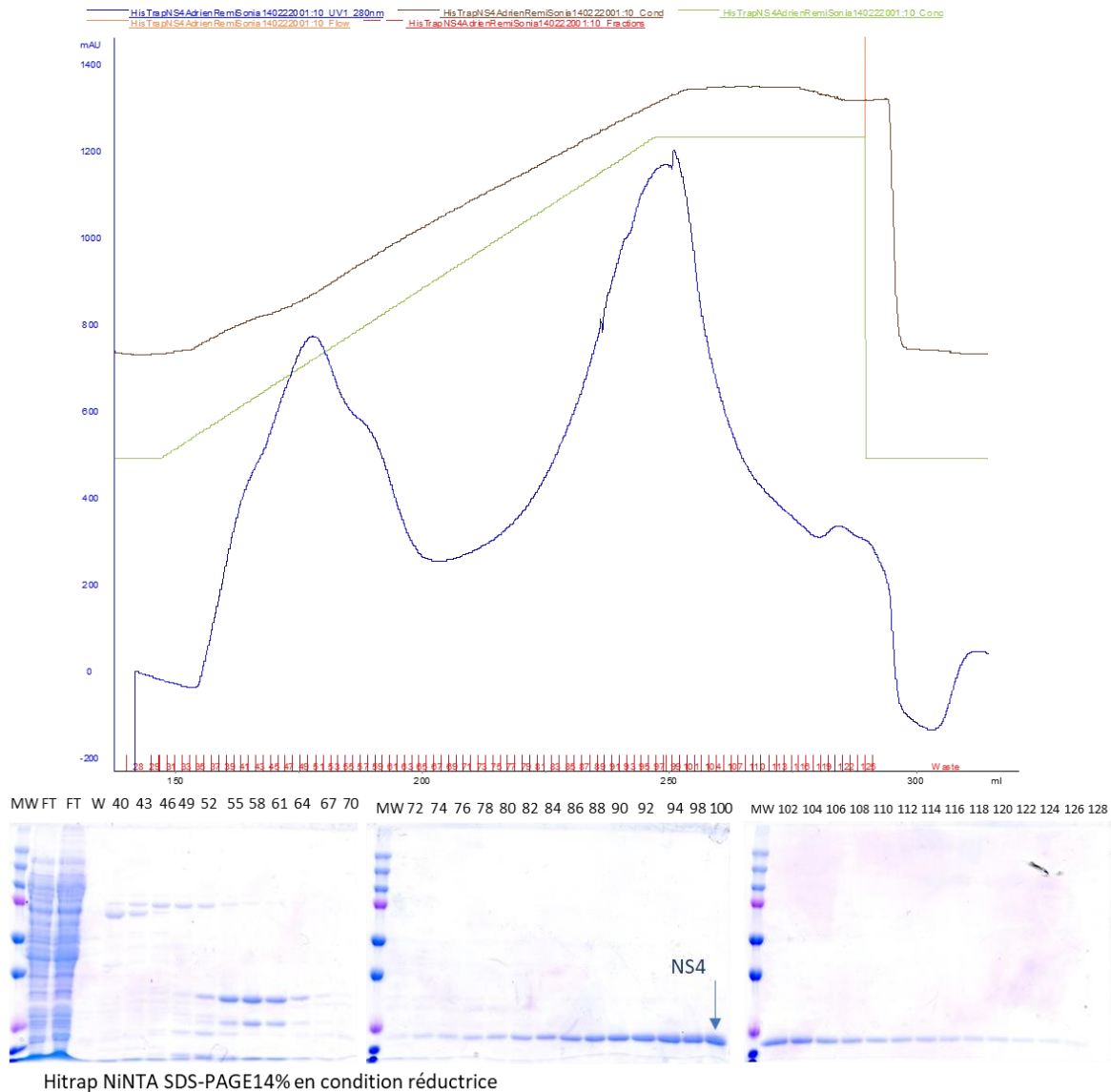
##### 3.1.1.1 Clonage et expression dans *Escherichia coli* BL21

Le protocole est identique pour NS4<sub>FL</sub>, NS4<sub>C160A</sub> et NS4<sub>tronqué</sub>. Le gène est inséré dans le vecteur d'expression bactérien pET16b incluant un site de clivage du facteur Xa du côté N-terminal ainsi que l'introduction de 10-histidines en N-terminal pour une purification par affinité au nickel (Ni-NTA). Le plasmide ainsi modifié est purifié puis séquencé avant transformation des bactéries compétentes. Les bactéries thermocompétentes NEBTurbo sont transformées avec 5 µl de plasmide à 100ng/ µl. Pour cela, les bactéries sont plongées dans la glace pendant 10min, puis 1 µl du vecteur est ajouté. Après 30min supplémentaires dans la glace, les bactéries sont incubées à 42 °C pendant environ 45 secondes. Elles sont à nouveau incubées dans la glace pendant 5min avant d'être reprises dans 900 µl de milieu de lysogénie (LB) puis incubées 1h à 37 °C sous agitation vive. Les bactéries sont ensuite centrifugées 5 min à 1000 g et le culot est solubilisé dans 100 µl de LB puis étalé sur une boîte de pétrie contenant du LB, de l'agar et de l'ampicilline puis incubé à 37 °C sur la nuit. Plusieurs clones sont utilisés pour créer une préculture de 300 ml de LB à 100 µg/ml ampicilline sur la nuit à 37 °C sous agitation vive. Le lendemain 5 litres de cultures sontensemencées à une densité optique (OD) de 0,15-0,2 mesurée à 600 nm. Une fois que l'OD a atteint ~0,8 la culture est induite avec 0,3mM d'IPTG pendant 5h à 37 °C. Les culots sont récupérés puis assemblés dans un Falcon de 50 ml après centrifugation durant 20min à 4000g puis stockés à -20 °C.

##### 3.1.1.2 Purification de NS4<sub>FL</sub>

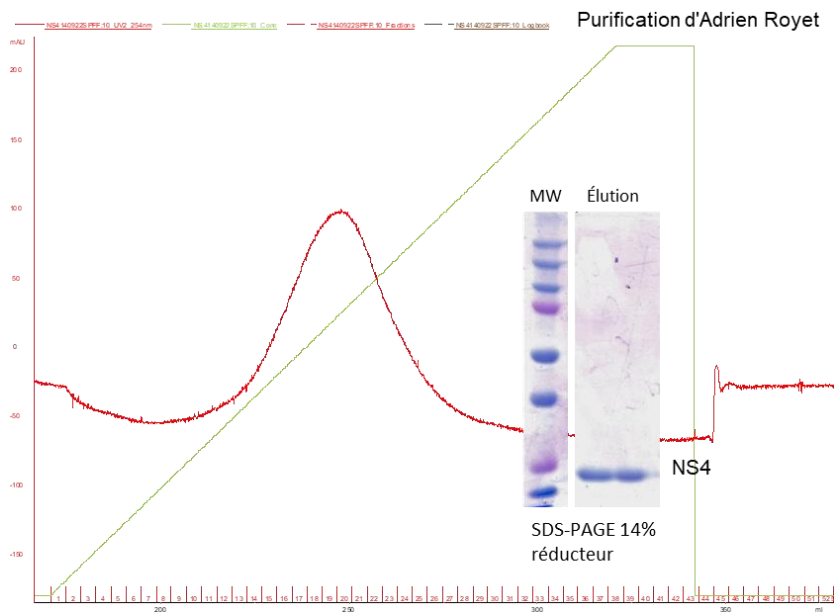
La purification de NS4 se déroule en trois étapes. Le culot est repris dans 80 ml/5L de tampon A (tableau 3.1.1.2-1) et incubé quelques minutes à 4 °C. Le volume est ensuite traité par sonication par 10 cycles composés de 30 secondes d'ultrasons et 1min de repos à 30 % de puissance, le tout dans la glace. Le produit de la sonication est centrifugé à 20000g durant 30min puis le surnageant est filtré à 0,22 µM. Le surnageant est injecté par pompage direct dans une colonne His-Trap (GE healthcare) de 5 ml préalablement équilibrée dans le tampon B. Une fois le surnageant intégralement injecté la colonne est lavée avec 50 ml de tampon B puis l'élution est effectuée avec un gradient de 0 à 100 % de tampon C sur 50 ml et un

fractionnement de 1,5 ml (figure 3.1.1.2-1). Les fractions sont contrôlées par un gel SDS-PAGE de 12 %.



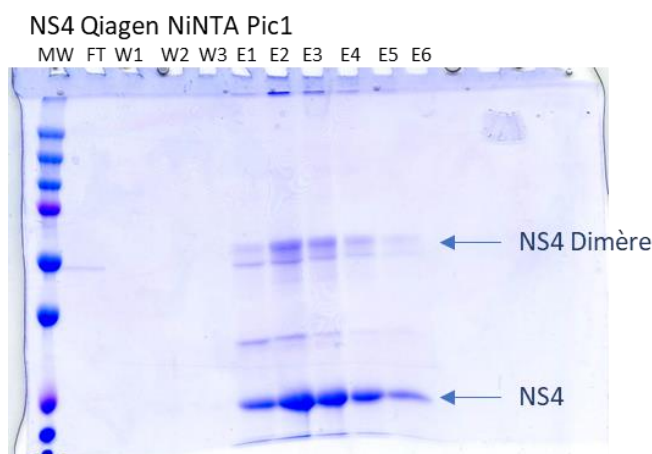
**Figure 3.1.1.2-1.** Résultat de l'étape de purification sur résine de nickel HiTrap de NS4<sub>FL</sub>. En bleu la courbe de l'absorbance UV à 280 nm, en vert le suivi du gradient, en marron la mesure effective de la conductimétrie.

Les fractions qui contiennent NS4<sub>FL</sub> sont changées de tampon par dilution d'un facteur deux dans le tampon D. Le volume est injecté sur la chromatographie échangeuse de cation SP-FF par pompage direct. Une fois l'échantillon chargé, la colonne est lavée avec 50 ml de tampon D. L'éluion est programmée avec un gradient de 0-100 % de tampon E sur 120 ml avec des fractions de 1,5 ml. Les fractions sont contrôlées par un gel SDS-PAGE de 15 %.



**Figure 3.1.1.2-2.** Exemple du profil d'élution lors de l'étape de purification sur SP-FF. Résultat obtenu par Adrien Royet. En rouge la courbe de l'absorbance UV à 245 nm. Le gel montre le profil typique de NS4 après cette étape sur gel SDS-PAGE 14% réducteur.

Les fractions dans lesquelles NS4<sub>FL</sub> a été identifiée sont incubées avec 1 ml de résine de Ni-NTA (Qiagen) préalablement lavée à l'eau et équilibrée dans un mélange à 50/50 % de tampon D et E. Le mélange résine et NS4<sub>FL</sub> est incubé 1h à 7 °C sous agitation par rotation douce dans un ou plusieurs tubes de 50 ml. Ensuite le tout est versé dans une colonne de paillasse avant de procéder à un lavage successif avec 5 ml des tampons F et G puis une élution dans 6 fractions de 500 µl de tampon H. Les fractions sont contrôlées par un gel SDS-PAGE de 15 %. Elles sont ensuite dialysées dans le tampon J avant congélation dans l'azote liquide pour être conservées à -80 °C. La même composition de tampon G sera utilisée pour toutes les expériences impliquant NS4.



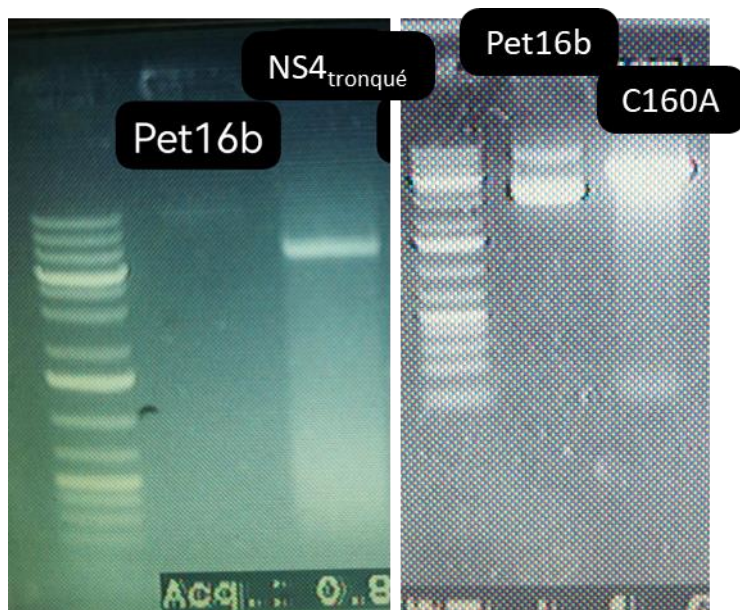
**Figure 3.1.1.2-3.** Résultat de l'élution de NS4 sur résine de Ni-NTA.

<b>Tampon A</b> Tris base 50mM pH7.4 NaCl 300mM Lysozyme 40mg/ ml EDTA free 2 pastilles Imidazole 20mM Triton X100 0,5 % Benzonase 55 µl β-OH 5mM	<b>Tampon B</b> Tris base pH7.4 50mm NaCl 300mM Imidazole 20mM Triton X100 0.02 % β-OH 5mM	<b>Tampon C</b> Tris base pH7.5 50mM NaCl 300mM Imidazole 50mM Triton X100 0.02 % DTT 1mM	<b>Tampon D</b> Tris base pH6.5 25mM NaCl 150mM Triton X100 0.1 % DTT 1mM	<b>Tampon E</b> Tris base pH6.5 25mM NaCl 1M Triton X100 0.1 % DTT 1mM
<b>Tampon F</b> MES pH6.5 10mM NaCl 150mM Triton X100 0,05 % DDM 0,04 % DTT 1mM	<b>Tampon G</b> MES pH6.5 10mM NaCl 150mM DDM 0,04 % DTT 1mM	<b>Tampon H</b> MES pH6.5 10mM NaCl 150mM DDM 0,04 % DTT 1mM Imidazole 500mM	<b>Tampon J</b> MES pH6.5 10mM NaCl 150mM TCEP 0.5mM	

**Tableau 3.1.1.2-1.** Résumé de la composition des différents tampons. β-OH = beta-mercaptoéthanol. DTT= Dithiothréitol, DDM= Dodécyl-β-D-maltoside, TCEP= Tris Carboxy Ethyl Phosphene.

### 3.1.1.3 Mutagénèse sur NS4

Les mutants NS4<sub>C160A</sub> et NS4<sub>tronqué</sub> sont créés à partir du gène de NS4<sub>FL</sub> par le protocole "quickchange II" de Agilent. Pour cela les amorces sont conçues sur le site internet Primer X puis commandées chez Eurofins. La PCR est lancée à partir de 50ng de plasmide pET16b contenant NS4<sub>FL</sub> dans un mélange réactionnel contenant 10 µM des amorces sens et antisens, 200 µM de nucléotide, 2 unités d'ADN polymérase phusion, du tampon spécial pour les régions riches en GC et 5 % DMSO. La PCR pour NS4<sub>tronqué</sub> est programmée comme suit : Une première dénaturation à 98 °C pendant 1min suivie de 16 tours de dénaturation à 98 °C pendant 30sec, d'hybridation des amorces à 68 °C pendant 30 secondes, et d'élongation à 72 °C pendant 3min30. La vitesse de synthèse de l'ADN polymérase est de 33 à 66 paires de bases par secondes et la taille de notre plasmide est d'environ 6270kb. Pour NS4<sub>C160A</sub> le programme est légèrement différent et commence par une première dénaturation à 98 °C pendant 1min, puis 16 tours composés d'une dénaturation à 98 °C pendant 30sec et d'une hybridation et élongation réalisées ensemble à 72 °C pendant 3min. L'ADN parental est digéré pendant 1h à 37 °C dans 5 µl de tampon CutSmart (NEB) avec l'enzyme DpnI (NEB). L'amplification est contrôlée par un gel d'agarose 10 % (0.6 g / 60 ml) avec 6 µl de colorant SyproSafe (figure 3.1.1.1-1).



**Figure 3.1.1.3-1.** Résultat du gel d'agarose après mutation et amplification des mutant NS4<sub>tronqué</sub> à gauche et NS4<sub>C160A</sub> à droite.

Nous travaillons à partir de la séquence NS4 entière génogroupe II génotype 4 variant *New Orleans* (*GII.4/NewOrleans*) suivante :

```
HHHHHHHHHHGPAITTFNFDRNKVLAFRQLAAENKYGLMDTMRVKGQLKDVRTMPELRQALKNVSIIK
KCQIVYSGCTYMLESDGKGNVKVDRIQSAAVQTNNELAGALHHLRCARIRYYVKCVQEALYSIIQIAGAAF
VTTRIAKRMNIQDLWSKPQVENTEEATSKDGCPCPKDDEELVISSDDIKTE
```

L'analyse de la séquence par l'outil ExPASy ProParam indique une masse moléculaire de 21502,56 g/mol et un point isoélectrique de 8,93. La séquence contient 5 cystéines et un seul tryptophane. Le coefficient d'extinction molaire en condition réductrice est de 14440 M<sup>-1</sup>.cm<sup>-1</sup> d'où une absorbance à 280 nm estimée à 0,67 pour 1 g/l.

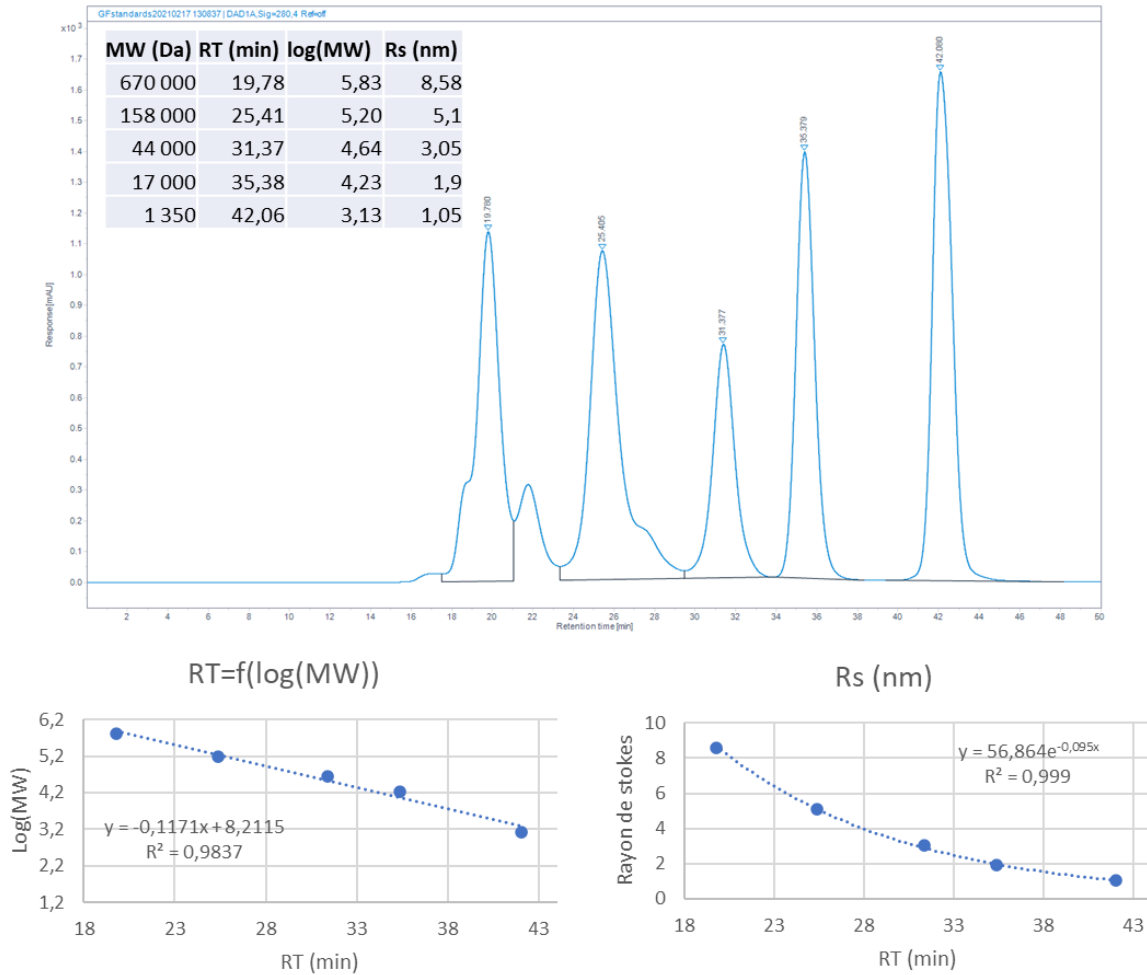
#### 3.1.1.4 Caractérisation biophysique

**Le spectre UV.** Un spectre UV est réalisé après chaque purification selon le protocole décrit dans le chapitre précédent.

**Prometheus Panta: DLS et Nano-DSF.** Les mesures de nano-DSF sont effectuées comme décrites dans le chapitre précédent. Excepté que les mesures DLS sont faites à 20 °C. Les mesures de nano-DSF sont faites sur une gamme de températures allant de 15 °C à 95 °C avec un pas de 1 °C. Les données sont traitées avec le logiciel nanotemper ou excel.

**SEC-SLS.** L'échantillon fraîchement purifié est analysé par SEC-SLS à l'institut Pasteur. Pour cela 50 µl d'échantillon est injecté et analysé dans un OMNISEC (Malvern Panalytical) avec une Superdex 200 10/300 GL couplé à un détecteur de diffusion statique de la lumière (SLS) et un réfractomètre. Le tampon d'éluion est composé de 10mM MES pH 6,5, 150mM NaCl 0.05 % DDM et 0.5mM TCEP.

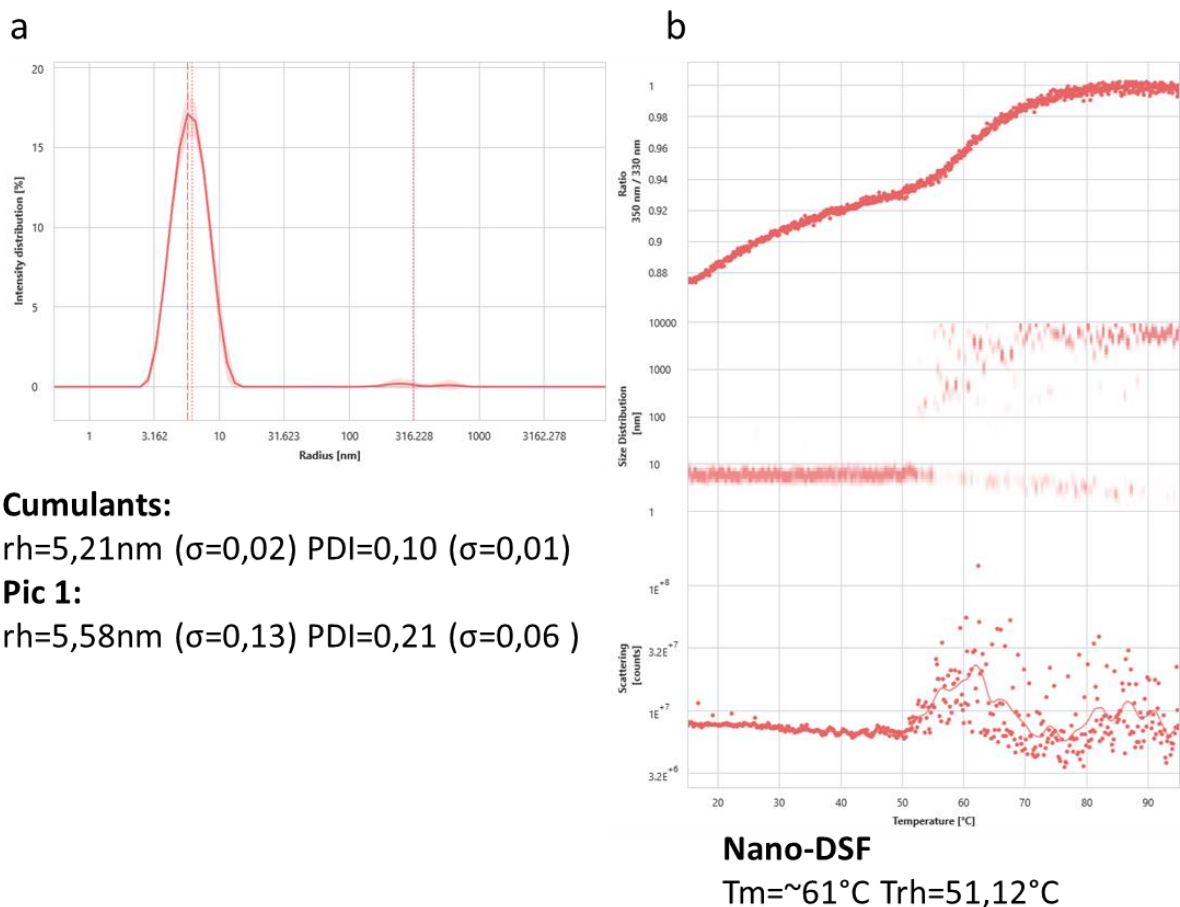
**FPLC calibration.** Pour réaliser la courbe de calibration, différentes protéines aux poids moléculaires connus provenant du kit de calibration distribué par Biorad sont injectées ensemble dans la S200 3,2/300. Les volumes de rétention sont mis en relation avec leurs masses moléculaires, de même que les rayons de Stoke (figure 3.1.1.4-1).



**Figure 3.1.1.4-1.** Profil de calibration obtenu avec le kit Biorad sur la S200 10/300 et les équations reliant le temps de rétention au poids moléculaire apparent ou au rayon de Stoke.

### 3.1.2 Résultats : Une vue exhaustive de NS4

Nous allons ici présenter les premières et uniques données biophysiques disponibles sur la protéine NS4. Commençons par les mesures nano-DSF et DLS. La DLS indique un rayon hydrodynamique de 5,21 nm ( $\pm 0.02$ nm) et un indice de polydispersité de 0,1 ( $\pm 0.01$ ). Le  $T_m$  de NS4 est d'environ 61 °C avec une agrégation visible à partir de  $\sim 51$  °C. Le profil des mesures DLS pendant le gradient de température (figure 3.1.2-1b size distribution) a la particularité de se scinder en deux populations (figure 3.1.2-1b). Nous pensons que les signaux associés aux gros objets sont provoqués par l'agrégation de NS4 tandis que la population du bas provient du signal du détergent. Ces deux mesures sont très stables et reproductibles d'une purification à l'autre.

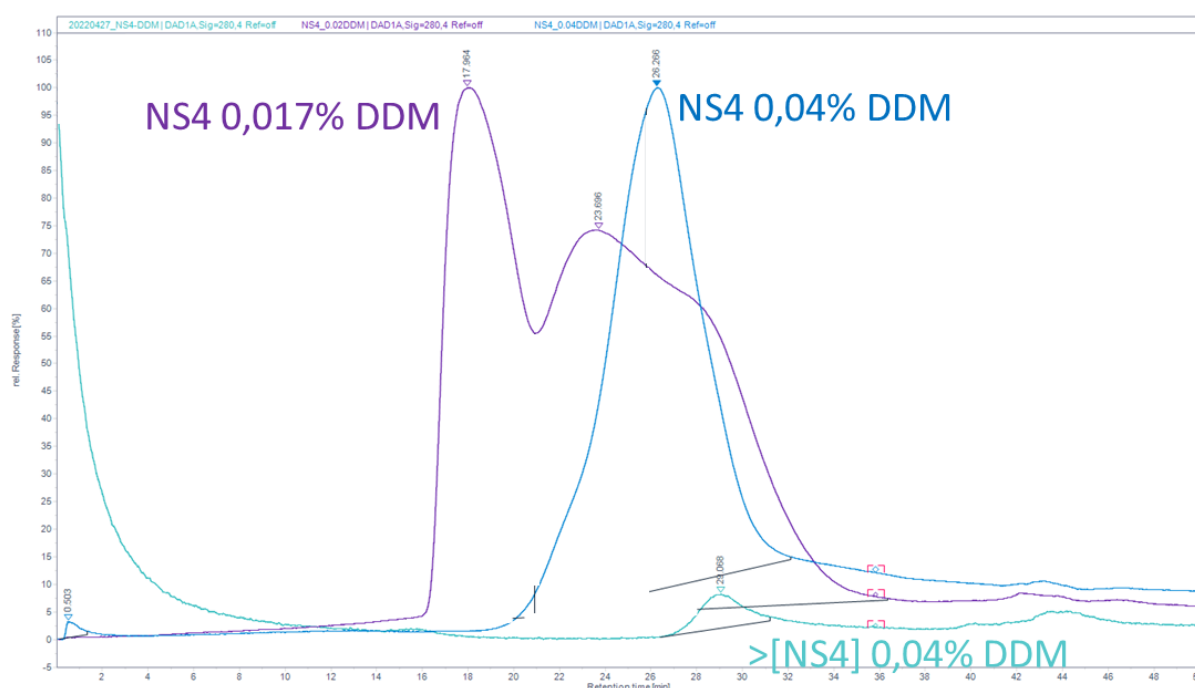


**Figure 3.1.2-1.** Résultats des mesures DLS (a) et nano-DSF (b) de la protéine NS4<sub>FL</sub> dans un tampon contenant 0,04 % de DDM.  $T_m$  = température de fusion,  $rh$  = rayon hydrodynamique,  $PDI$  = indice de polydispersité.

NS4 est une protéine membranaire monotopique. Les travaux précédents réalisés au sein de l'équipe et par des collaborateurs ont apporté des connaissances de base sur le comportement de NS4<sub>FL</sub>. En revanche il n'existe aucune donnée structurale sur la protéine. Johann Habersetzer, lors d'un postdoc réalisé précédemment, a obtenu des résultats préliminaires révélant un changement d'état d'agrégation en fonction de la concentration de



NS4 et de la concentration de détergent (résultats internes). À partir de ces résultats, différentes injections en FPLC de NS4 à différentes conditions ont été réalisées (figure 3.1.2-1). La même quantité de NS4 est injectée en FPLC dans 3 concentrations de détergent différentes et sans agent réducteur, la première dans 0,04 % de DDM, la seconde à 0,02 % de DDM et la dernière sans DDM. À 0,04 % de DDM le signal d'éluion est majoritairement à 26,26min avec deux petits épaulements mineurs avant et après le centre du pic. La diminution du DDM à 0,02 % entraîne un changement du profil d'éluion avec l'apparition d'un pic majoritaire à 17,9min correspondant à un agrégat, un second pic à 23,7min avec un épaulement majeur à environ 28,5min. L'injection de NS4 sans détergent entraîne un signal exclusivement en agrégat et fortement diminué. De même la diminution de la quantité de NS4 augmente son temps de rétention (figure 3.1.2-1 et tableau 3.1.2-1).



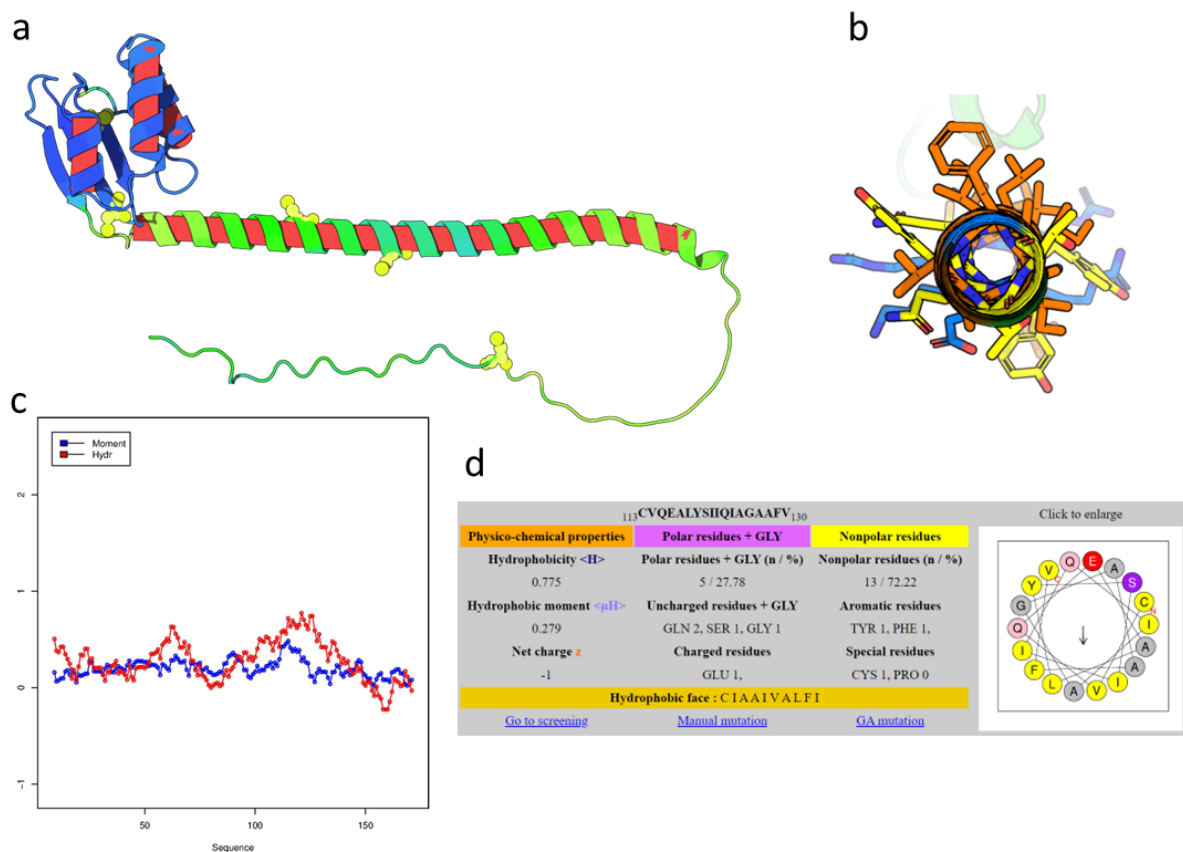
**Figure 3.1.2-1.** Résultats des injections FPLC de NS4<sub>FL</sub> en différentes concentrations de détergent DDM et en différentes concentrations de NS4.

Concentration de détergent	Quantité injectée	Temps de rétention (min)	Masse moléculaire apparente (rStokes)
0,04 % DDM	NS4 [56 µM / 10 µl / 0,56 nmol]	29,07	73kDa (3,6)
0,017 % DDM	NS4 [56 µM / 50 µl / 2,8 nmol]	17,9 (Aggregate) + 23,7 + 28,5	>255kDa + 83kDa (3,8)
0,04 % DDM	NS4 [56 µM / 40 µl / 2,2 nmol]	26,3	139kDa (4,6)

**Tableau 3.1.2-1.** Concentration en DDM et en NS4 aux différentes injections. Temps de rétention et poids moléculaire apparent ainsi que le rayon de Stokes selon la courbe de calibration.

Les résultats précédents montrent un lien étroit entre la protéine et le détergent sans lequel

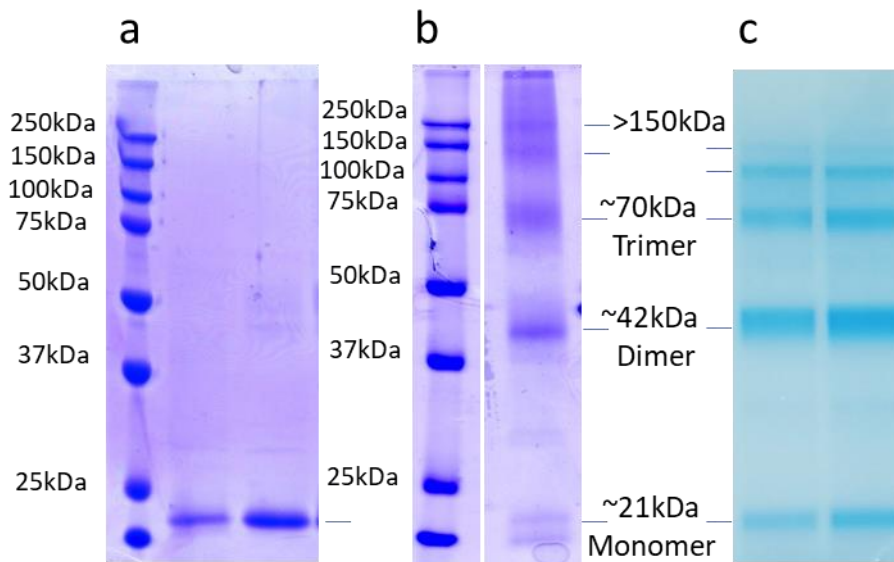
elle ne peut être purifiée correctement. Par ailleurs, les analyses bio-informatiques à partir de la séquence de NS4 permettent de mieux comprendre et appréhender son comportement. Dans un premier temps les prédictions alphafold convergent vers un repliement qui peut être défini en trois domaines : un domaine globulaire soluble, une longue hélice suivie d'une partie c-terminale non structurée (figure 3.1.2-2a). Cette organisation confirme les prédictions réalisées précédemment dans l'équipe (Doerflinger 2017). La longue hélice n'a pas de propriétés transmembranaires mais semble plutôt posséder les caractéristiques d'une protéine membranaire monotopique, avec le bout c-ter de l'hélice présentant une face clairement apolaire (acides aminés en orange) et l'autre chargée (figure 3.1.2.2b). Sur les graphiques obtenus par Heliquist, nous observons un moment hydrophobe autour des acides aminés 113-130 qui correspond à la zone monotopique (figure 3.1.2-2c et 2d).



**Figure 3.1.2-2.** Prédiction alphafold de NS4<sub>FL</sub> (a) avec les cystéines représentées en modèle atomique colorées en jaune et la protéine colorée par le score d'indice de confiance d'alphafold du plus haut score en bleu vers le plus bas en vert. A droite (b) du modèle la vue transverse de la longue hélice avec les acides aminés colorés en orange pour les non-polaires, en jaune pour les polaires non chargés et en bleu pour les polaires chargés. En dessous les résultats de heliquist avec le graphique du moment hydrophobe en fonction de la séquence (c) et à droite l'analyse détaillée centrée autour des acides aminés 113-130 (d).

Lors de ces premières expériences, la protéine était manipulée sans agent réducteur à partir

de la dernière étape sur gel filtration. Or la protéine contient 5 cystéines, et les prédictions alphafold révèlent que 3 d'entre elles sont accessibles. À partir de cette observation, nous avons réalisé une caractérisation rapide sur gel SDS-PAGE dénaturant mais non réduit en comparaison à un gel dénaturant et réduit. Cette expérience révèle la présence de bandes distinctes même à très haut poids moléculaire suggérant la présence de forme multimérique de NS4 impliquant les cystéines (figure 3.1.2-3).

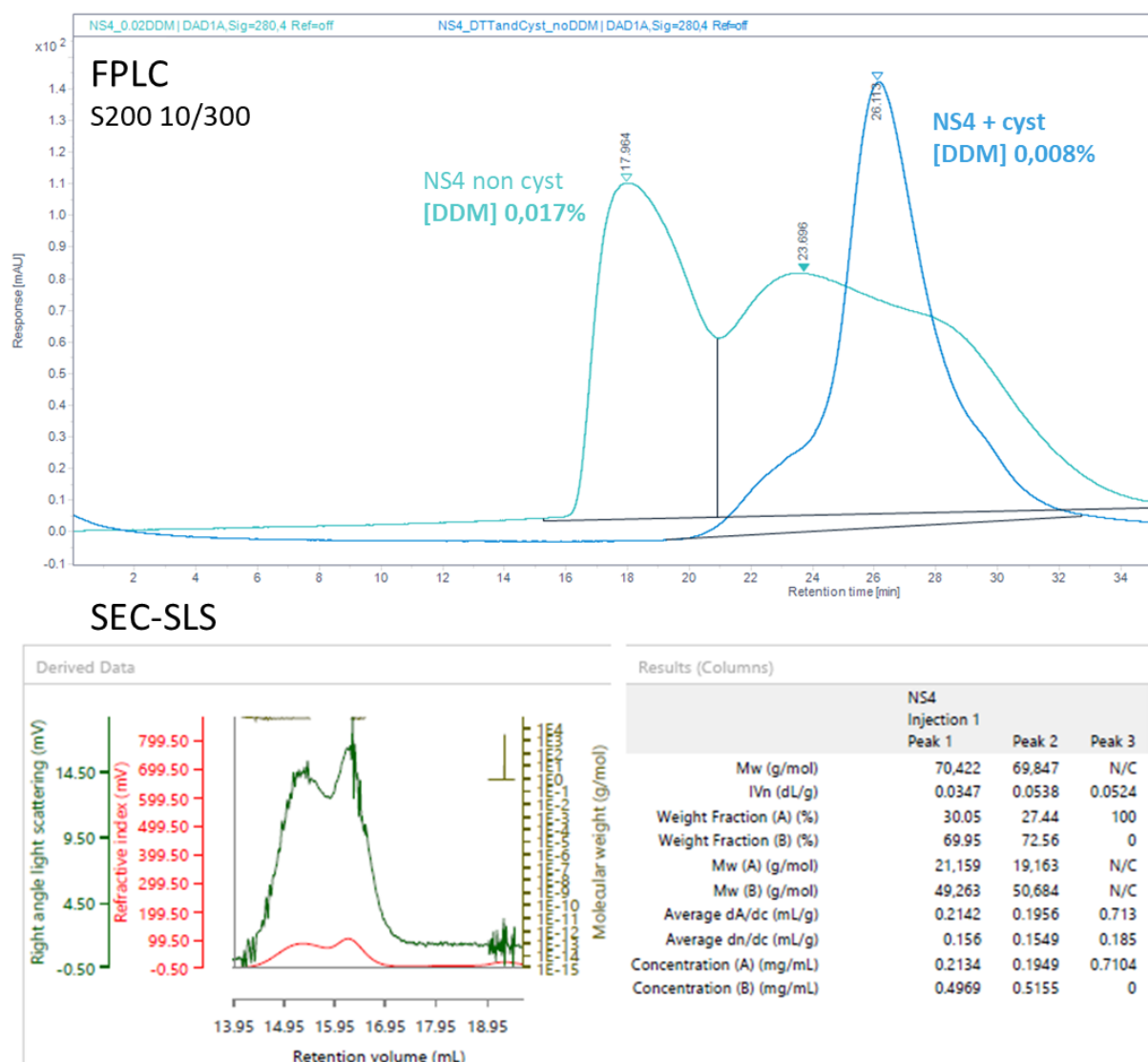


**Figure 3.1.2-3.** Comparaison des gels SDS-PAGE de la protéine NS4<sub>FL</sub> en conditions réductrices et non réductrices. En **a** le gel SDS-PAGE 15 % en condition réductrice et dénaturante. En **b** le même échantillon de NS4<sub>FL</sub> sans agent réducteur durant la migration sur gel. En **c** même chose que **b** mais en utilisant un gel en gradient d'acrylamide commercial plus résolutif.

Basé sur ces observations, nous avons réalisé de nouvelles injections en FPLC en présence d'agent réducteur ou stabilisant (DTT, TCEP ou Cystéine). Dans l'idée de travailler plus tard avec des Fabs dont le pont disulfure entre la chaîne lourde et légère est un élément crucial à son assemblage et sa stabilité, nous avons testé la stabilité de NS4 en présence de cystéine dans le tampon d'élution. Pour cela NS4 est préalablement incubé dans 5mM de DTT puis injecté en FPLC dans un tampon contenant 20mM de Cystéine mais 0,008 % de DDM c'est-à-dire proche de sa CMC. Là où précédemment nous avons un changement radical du profil d'élution dès la diminution du DDM à 2xCMC, nous observons ici un profil d'élution très proche des conditions à plus haute concentration de DDM avec un temps de rétention de 26,1min soit un poids moléculaire apparent de 146kDa et un rayon de Stokes de 4,8 (figure 3.1.2-4 et tableau 3.1.2-2). Une observation similaire a pu être faite sur SEC-SLS à Pasteur. Dans cette expérience nous avons injecté la protéine NS4 en présence de TCEP, un puissant agent réducteur, et de 0,02 % de DDM. Le profil d'élution montre deux pics proches difficiles à interpréter. Le poids moléculaire mesuré de NS4 est de 70kDa, cohérent avec les estimations précédemment faites en FPLC. À partir des données de réfractométrie nous pouvons identifier deux espèces présentes dans les pics. Les données convergent vers un ratio proche des

30 %/70 % entre la protéine NS4 et le détergent DDM. Il en résulte une extrapolation du poids moléculaire d'environ 19 ou 21kDa pour le composé A (NS4) et ~50kDa pour le composé B (micelle de DDM). Cela est cohérent avec la littérature qui rapporte également des micelles de DDM proches des 50kDa (Lipfert et al. 2007; Hutchison et al. 2017).

Ces observations montrent qu'en plus d'être sensible au ratio de la concentration de DDM, la protéine NS4 est aussi sensible aux conditions rédox (Figure 3.1.2-4 et tableau 3.1.2-2). Elles montrent aussi qu'en fortes conditions réductrices, il semble possible de stabiliser cette protéine sous une forme homogène qui semble être un monomère. Les comparaisons des profils d'élutions avec et sans cystéine dans le tampon montre que la formation de gros objets en solution est directement liée à la formation d'un pontage impliquant les cystéines.



**Figure 3.1.2-4.** En haut, comparaison en FPLC des profils d'élution de NS4<sub>FL</sub> injection en présence (courbe bleue) et en absence (courbe cyan) de cystéine dans le tampon d'élution. En bas, résultats des mesures en SEC-SLS. Le tableau indique les mesures effectuées sur les pics d'élution visibles en rouge. Les mesures sont : le poids moléculaire (MW), le temps de rétention

(RV), le signal associé au pic (Recovery), la part de l'échantillon dans le pic (Frac. of sample), et la concentration dans la fraction du pic (Peak measured conc.).

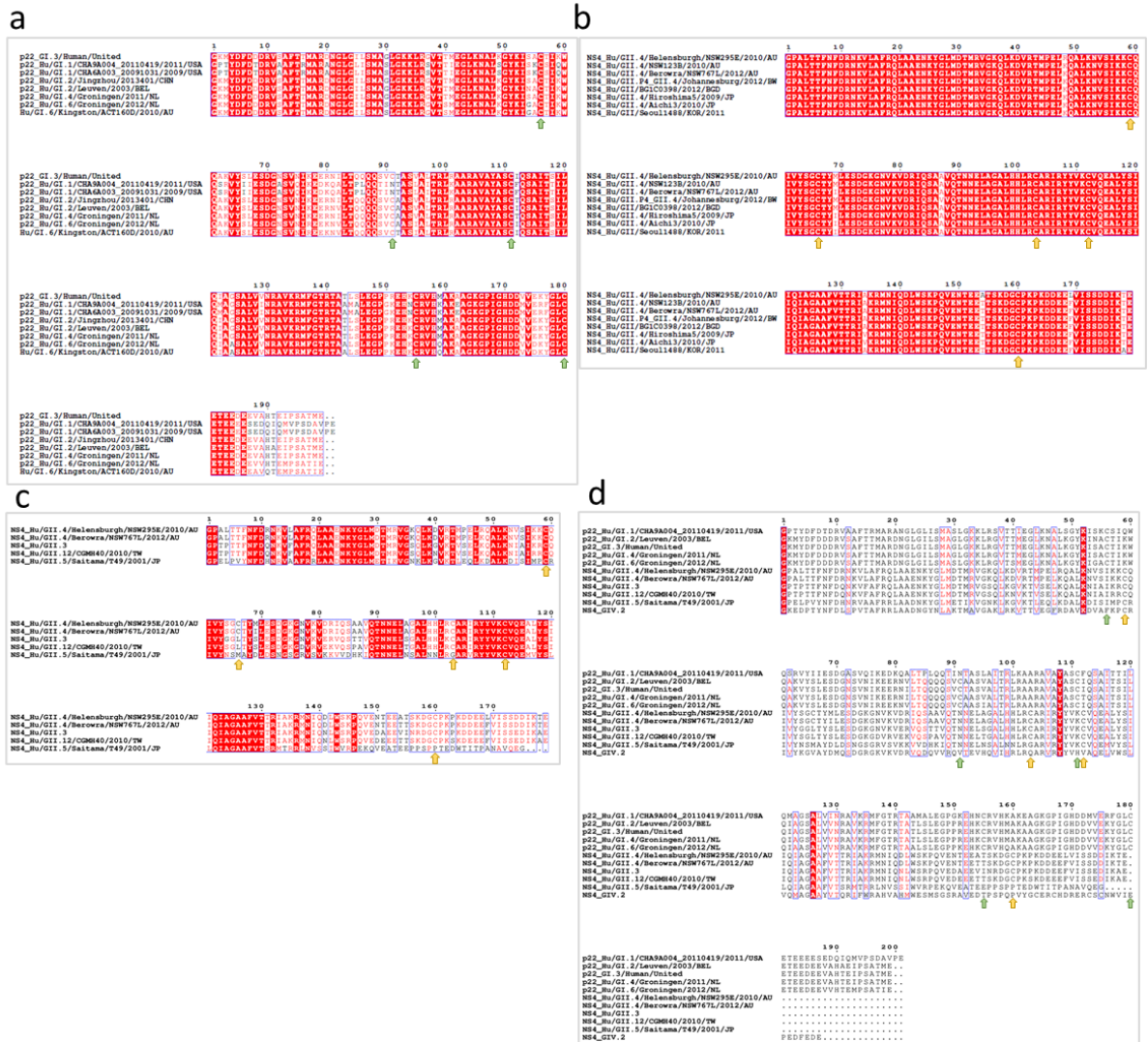
<b>Composition du tampon en détergent</b>	<b>Quantité de protéine injecté</b>	<b>Temps de rétention (min)</b>	<b>Masse moléculaire apparente (rayon de Stoke)</b>
0,04 % DDM (sans cystéine)	NS4 [96 $\mu$ M / 30 $\mu$ L]	27,05	105kDa (4,1)
0,008 % DDM + 2 mM cystéine	NS4 [96 $\mu$ M / 30 $\mu$ L]	26,1	146kDa (4,6)

**Tableau 3.1.2-2.** Résumé des injections de NS4 avec et sans présence de Cystéine dans le tampon.

Nous venons de montrer le rôle des cystéines dans le pontage intermoléculaire en condition biochimique. Nous allons terminer la réflexion sur NS4 par la comparaison des séquences des génogroupes I, II et IV qui sont les seuls qui infectent l'homme.

Les alignements des séquences de NS4 du génogroupe I génotypes 1, 2, 3, 4 et 6 montre que cinq cystéines (Cys56, Cys91, Cys111, Cys155, Cys180) sont strictement conservées au sein du génogroupe excepté chez le génotype 1 (GI.1) où la cystéine 91 est mutée en asparagine 91 (figure 3.1.2-5a). Dans le génogroupe II, les génotype 4 partagent la même séquence (figure 3.1.2-5b). En revanche, les séquences divergent un peu entre les génotypes 3, 4, et 12 (GII.3, GII.4, GII.12). Dans ce cas, cinq cystéines sont conservées (Cys59, Cys66, Cys103, Cys112, Cys160) entre GII.3 et GII.4 mais seules les Cys59 et Cys112 sont conservées avec GII.5 (figure 3.1.2-5c).

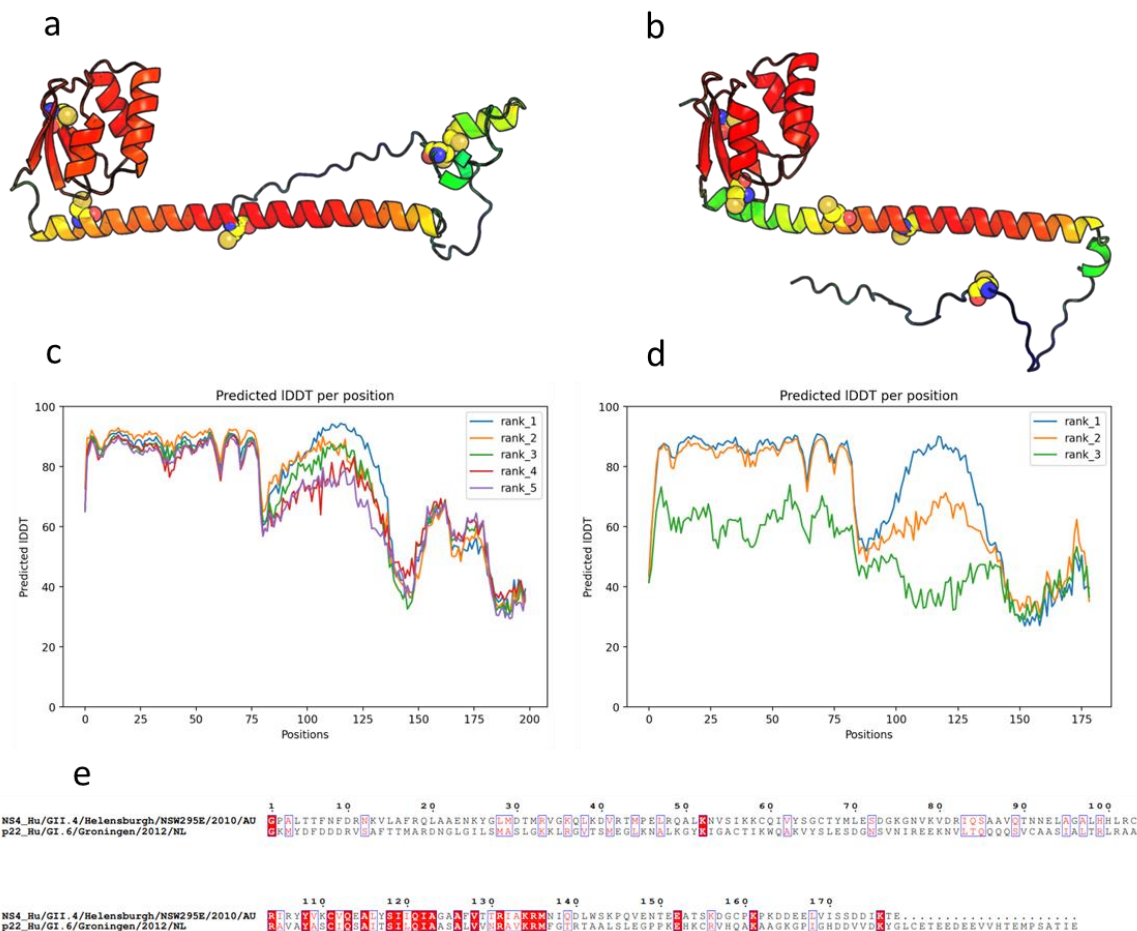
En revanche, d'un génogroupe à l'autre (entre GI-1,2,3,4,6, GII-3,4,5,12 et GIV) les cystéines ne sont pas rigoureusement conservées. Entre GI et GII la plupart d'entre elles sont toujours proches dans la séquence, éloignées de 2 à 5 acides aminés, sauf pour la Cys66 de GII qui ne semble pas avoir d'équivalent dans GI (figure 3.1.2-5c). Pour GIV le placement des cystéines complètement différent sauf pour la Cys59, mais qui est placé dans le domaine globulaire. En revanche nous observons que 4 acides aminés sont rigoureusement conservés (Glycine 1, Lysine 52, Tyrosine 108, et Alanine 106), et d'autres régions s'éloignent très peu d'un point de vue des propriétés physico-chimiques des acides-aminés.



**Figure 3.1.2-5.** Alignements de séquences de NS4 issues des génogroupes I et II. Le panneau **a** montre l'alignement des génotypes 1, 2, 4 et 6 du génogroupe I. Le panneau **b** montre l'alignement de différents variants de GII.4. Le **c** montre l'alignement des génotypes 3, 4, 5, et 12 du génogroupe II. Enfin le panneau **d** montre les alignements de GI.1-2-4-6 avec GII.3-4-5-12 et GIV.2 en dernière ligne. Les cystéines du GI sont indiquées par des flèches vertes et celles du GII par des flèches jaunes. Les acides aminés sur fond rouge sont ceux strictement identiques et ceux écrits en lettre rouge ou encadré en bleu ont des propriétés physico-chimiques proches. (Alignements disponibles en haute définition en annexes).

Nous avons réalisé des modèles alphafold à partir des séquences de la protéine NS4 de GII.4/Helensbourg/2010 et GI.6/Groningen/2012 (figure 3.1.2-6). Bien qu'étant des protéines homologues, elles ne sont liées que par un faible 33,14 % d'identité de séquence. Malgré cette faible similitude les modèles alphafold de ces deux génogroupes révèlent un repliement très proche (globulaire, longue hélice, domaine non-structuré). Elles sont toutes deux prédites avec une cystéine située au centre de l'hélice amphipatique. En revanche, dans la zone qui suit la longue hélice, chez GI.6 (figure 3.1.2-6a), les deux cystéines (155 et 180) sont prédites très proches à l'interface entre deux hélices avec un haut score de confiance. Elles seraient

donc, dans milieu non-réducteur, plutôt impliquées dans un pont disulfure intramoléculaire. Pour GII.4, dans la même région, il y a une seule cystéine unique (figure 3.1.2-6), qui pourrait être responsable du pontage par des ponts disulfures intermoléculaire observés sur les gels non-réduit de la figure 3.1.2-3. Le score de fiabilité LDDT (défini en partie 2.3.2.3) est très bon pour le domaine globulaire, ils sont légèrement moins homogènes mais reste acceptable le long de l'hélice amphipatique, et chutent pour le domaine non-structuré. Ces modèles révèlent la divergence du positionnement des cystéines dans la structure et est un élément à la défaveur de leur hypothétique rôle biologique.



**Figure 3.1.2-6.** Modèles alphafold de rang 1 issus de la séquence NS4 GI.6 à gauche (a) et NS4 GII.4 à droite (b) coloré par score de prédiction du plus haut indice de confiance au plus faible du rouge au bleu respectivement. Les cystéines sont montrées en sphères jaunes. En dessous en c et d les graphiques du score LDDT par résidu pour chacun des rangs. Le test de différence de distance locale (IDDT) est un score qui évalue les différences de distance locale entre tous les atomes d'un modèle et reflète est bien adapté à l'évaluation de la qualité des modèles locaux. En bas l'alignement des deux séquences utilisées pour réaliser les modèles. Les acides aminés en rouge sont strictement identiques ceux encadrés en bleu ont des propriétés physico-chimiques proches.

### 3.1.3 Conclusion et discussion : une meilleure compréhension de NS4 par la biophysique et les outils bioinformatiques

La caractérisation biophysique de NS4 montre clairement l'implication des cystéines dans l'oligomérisation de NS4 GII.4 dans des conditions biochimiques, c'est-à-dire pour un échantillon purifié et concentré, et donc relativement éloignées des conditions biologiques. L'analyse de séquences indique une certaine forme de conservation au sein des deux principaux génogroupes Humain dans la structure et certaines régions des séquences. Mais les signaux de conservation des cystéines sont très faibles, ce qui laisse perplexe quant à leur rôle biologique. De plus, NS4 est exclusivement dans le cytosol qui est fortement réducteur. Bien qu'elle interagisse avec le réticulum endoplasmique (RE) après le recrutement de NS1-2 (Doerflinger et al. 2017; Hung et al. 2023) et bien que la lumière du RE soit un environnement qui permette la formation de ponts disulfures (Robinson and Bulleid 2020), aucune donnée ne montrent une internalisation de NS4.

Il existe une dernière observation qui discrédite l'hypothèse d'un rôle des cystéines dans le processus infectieux. Chez les poliovirus la protéine 3A jouent un rôle très proche de celui de NS4. La protéine 3A interagit également avec le RE (Teterina et al. 2011) et peut également déformer les liposomes (Wang et al. 2013) de façon similaire à ce que fait NS4 (travaux internes non publiés). De même, la protéine 3AB de poliovirus a un repliement (alphafold) similaire à celui de NS4 en se partageant en trois domaines : un domaine globulaire, une hélice amphipatique et un domaine déstructuré. Les cystéines ne seraient donc pas à l'origine de l'activité de remodelage membranaire et l'oligomérisation observé serait un pur artefact biochimique.

En revanche, la compréhension de comportement de NS4 lors de sa purification est primordiale pour optimiser les approches d'études structurales. D'un point de vue purement biochimique donc, le modèle alphafold de NS4 GII.4 nous permet d'imaginer un scénario dans lequel la protéine pourrait se ponter par les cystéines libres. La cystéine 160 créerait un pont disulfure avec une des cystéines de l'hélice amphipatique et pourrait ainsi former un réseau en théorie infini. C'est ce qui est observé sur les gels SDS-PAGE non réduits (figure 3.1.2.3-1). Nous savons à présent que la protéine NS4 a une tendance à l'oligomérisation qui évolue selon trois éléments. Le facteur principal est l'état réducteur du tampon qui va directement induire une oligomérisation de NS4 à mesure qu'elle s'oxyde. Le deuxième est la concentration de DDM qui, en diminuant, va favoriser l'apparition d'oligomère. Enfin, le troisième est la concentration de NS4 qui sera d'autant plus favorable à la formation d'oligomère qu'elle sera élevée.

Les travaux de thèse d'Adrien Royet et les tentatives précédentes de l'équipe indiquent que la protéine NS4 est difficile à cristalliser et difficile à étudier par RMN. Notamment à cause de ses propriétés hydrophobes et des pontages intermoléculaires par pont disulfures. L'étude structurale de la protéine NS4<sub>FL</sub> par cryo-EM est rendue complexe par sa petite taille (21kDa). Pour s'affranchir de cette contrainte nous avons tenté une approche similaire à celle



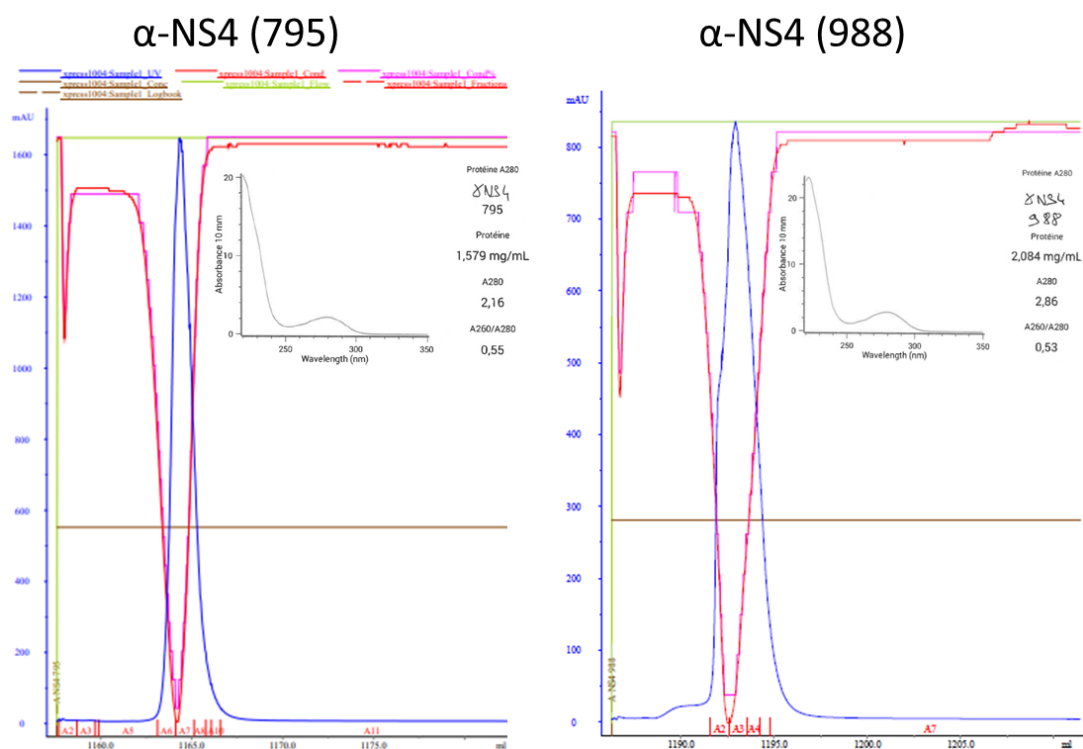
présentée pour le complexe HTP, c'est-à-dire fixer un ou plusieurs Fab(s) à NS4. L'objectif est de grossir artificiellement la protéine NS4 pour obtenir un objet significativement plus gros et plus adapté à une étude par cryo-EM.

## 3.2 Des Fabs pour une structure par cryo-EM de la protéine NS4

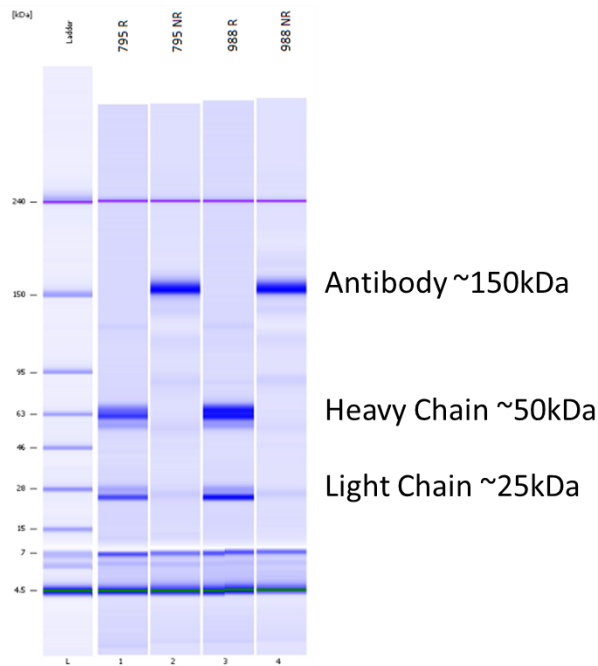
### 3.2.1 Matériel et méthode : purification des mABs et des Fabs

#### 3.2.1.1 Production et purification des anticorps monoclonaux

Les anticorps monoclonaux sont produits par des collaborateurs de l'équipe de Volker Lohmann du laboratoire de virologie moléculaire d'Heidelberg du centre de recherche Allemand "DKFZ". Ce sont des anticorps de souris de type IgG1 qui avaient été obtenus à partir de l'une des premières purifications de NS4 réalisées par Johann Habersetzer. Deux lignées monoclonales nommées 795 et 988 sont purifiées avec l'aide du laboratoire NG biotech au CEA via une colonne de type Hitrap protéine A de 5 ml Le chargement et le lavage sont faits dans un tampon de 50mM phosphate pH 7,5 pendant 5 volumes de colonne avant l'élution dans un tampon de 100mM glycine pH 2,5 (figure 3.2.1.1-1). Les fractions d'élution sont rassemblées et dialysées sur la nuit à 7 °C dans un tampon de 150mM NaCl et 50mM de phosphate de potassium à pH 7,5 puis le volume collecté après dialyse est analysé par UV sur un nanodrop. Les fractions après dialyse sont analysées par un gel SDS-PAGE sur un système Agilent microarray system (figure 2.1.1-2).



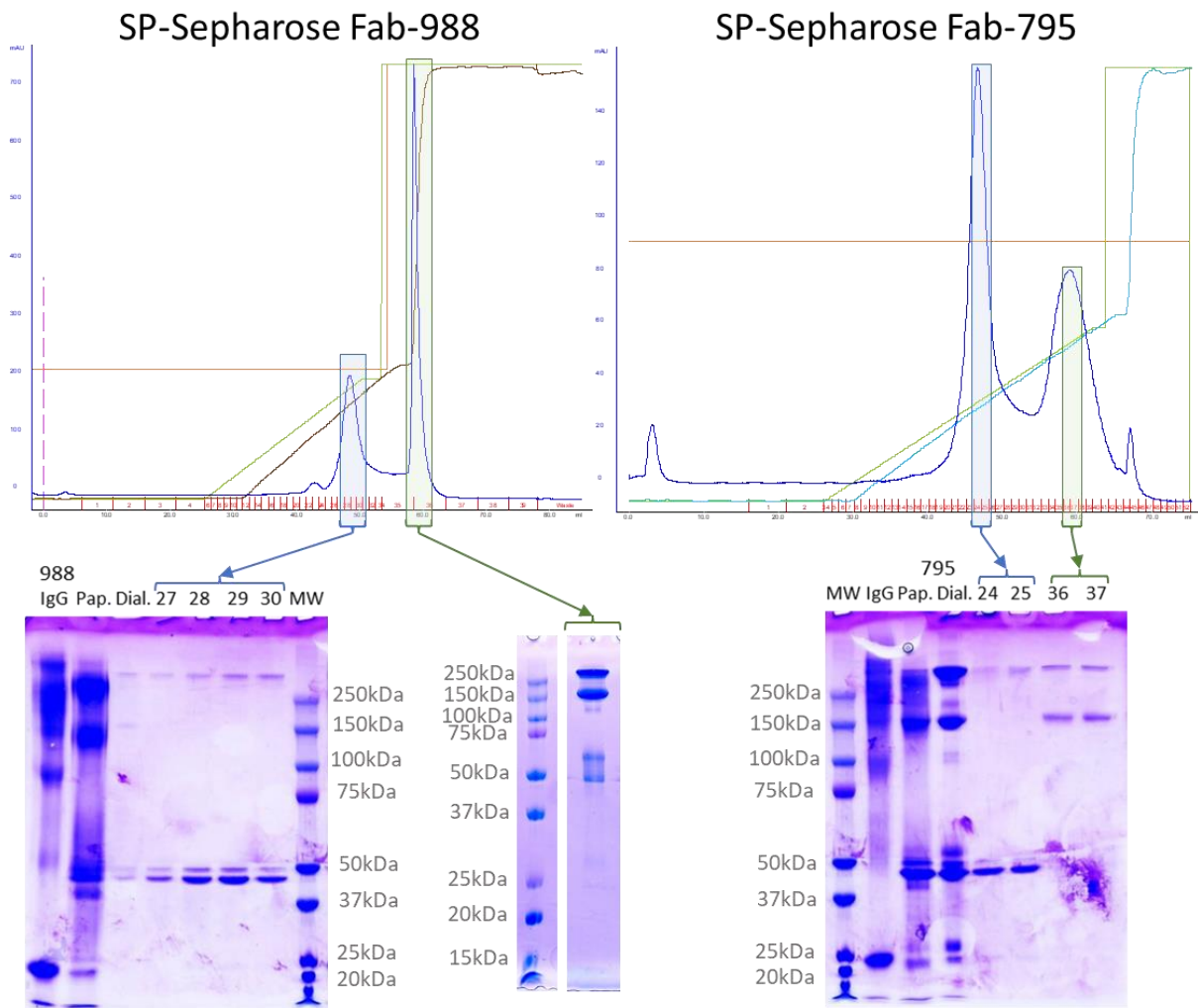
**Figure 3.2.1.1-1.** Profil d'élution des anticorps monoclonaux 795 et 988 avec le spectre UV et la mesure de leur concentration respective. En bleu la courbe de l'absorbance à 280 nm en rouge la conductivité, en vert le débit.



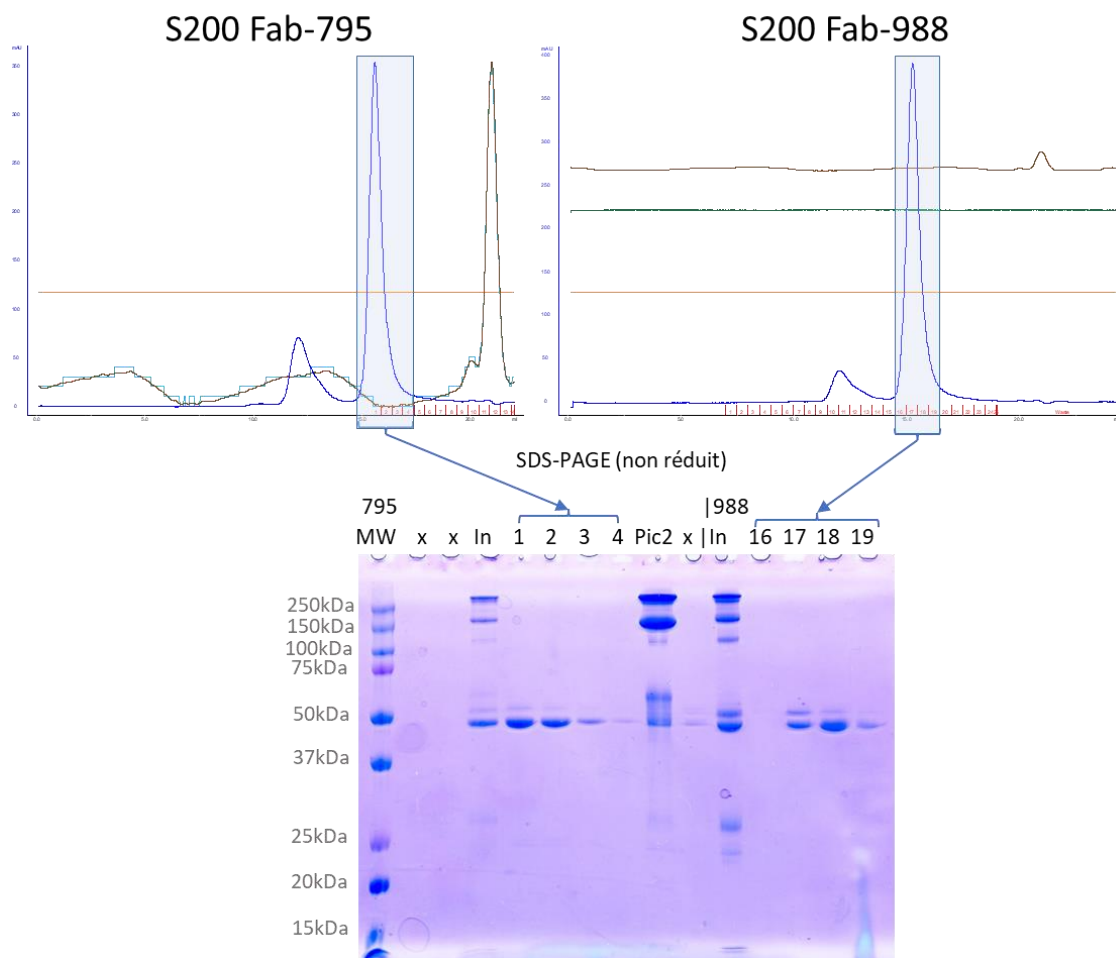
**Figure 3.2.1.1-2.** Gel SDS-PAGE 12 % obtenu par le système Agilent microarray. R: conditions réductrices ; NR : conditions non réductrices.

### 3.2.1.2 Production et purification des Fabs

Les Fabs sont obtenus de manière très similaire à la méthode décrite dans le chapitre précédent. Brièvement, les anticorps sont concentrés à 10 mg/ml puis après ajout de 20 mM de cystéine ils sont incubés à 37 °C sous agitation vive pendant 4 heures avec de la papaïne immobilisée sur résine préalablement lavée avec un tampon composé de 20 mM de phosphate de sodium pH7, 10 mM EDTA, 20 mM de cystéine. Une fois digérés, les Fabs sont équilibrés dans un tampon acétate à pH4 par centrifugation ou dialyse avant d'être purifiés par SP sepharose. L'équilibration, l'injection et le lavage sont faits dans 50 mM de tampon acétate pH4 tandis que l'élution est faite à 0,3 ml/min dans un gradient de 120 min s'étalant de 0 à 40 % du tampon composé de 50mM d'acétate et 1 M NaCl. Les fractions éluées sont neutralisées avec 100 µl de tampon phosphate pH 7 à 1 M. Les fractions sont analysées sur gel SDS-PAGE 14%. Les fractions avec les Fab sont dialysées sur la nuit à 7 °C dans 1L de tampon à 10 mM MES pH6,5 et 150 mM NaCl (Figure 3.2.1.2-1). Les Fabs sont ensuite concentrés et injectés en S200. Les fractions contenant les Fabs (Figure 3.2.1.2-2) sont concentrées puis fractionnées en aliquotes avant d'être congelées dans l'azote liquide et stockés dans le congélateur à -20 °C.



**Figure 3.2.1.2-1.** Profil et gel d'élution SP-Sepharose des Fabs. Les courbes bleues sont l'absorbance à 280 nm, en vert le gradient en pourcentage de tampon B, en marron et cyan la mesure de la concentration de sel (conductivité). En orange le débit. Les gels SDS-PAGE sont en condition non réductrice. IgG = Immunoglobuline avant la digestion par la papaine. Pap. = Echantillon après la digestion par la Papaine. Dial. = Échantillon digéré après dialyse. MW = Marqueur de poids moléculaire\*



**Figure 3.2.1.2-2.** Profil d'élution des Fab sur S200. Les courbes bleues sont l'absorbance à 280 nm, en marron la mesure de la concentration de sel (conductivité) et en orange le débit. Le Gel SDS-PAGE est en condition non réductrice et 14% de polyacrylamide.

### 3.2.1.3 Biophysique nano-DSF, DLS

Les mesures de nano-DSF et DLS sont réalisées comme expliqué dans le chapitre précédent. 10 µl d'échantillon est introduit dans un capillaire et analysé sur l'appareil prometheus Panta de chez Nano-temper.

### 3.2.1.4 MALDI-TOF

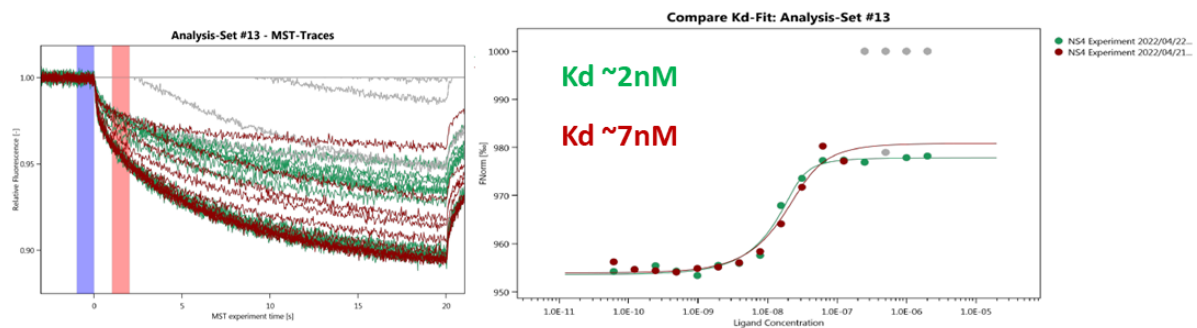
La spectrométrie de masse a été réalisée à l'institut Pasteur sur un spectromètre Bruker Daltonics - ultraflex TOF/TOF. Les Fabs sont mélangés à un solvant contenant de l'acétonitrile et du TFA. Une goutte de 0,5 µl est déposée sur la plaque ainsi que 1 µl de solution de matrice. L'échantillon est laissé séché avant de lancer l'analyse.

### 3.2.1.5 SEC-SLS

Le Fab-795 est analysé par SEC-SLS à l'institut Pasteur. Pour cela 50 µl d'échantillon sont injectés et analysés avec un OMNISEC (Malvern Panalytical) avec une Superdex 200 10/300 GL couplé à un détecteur de diffusion statique de la lumière (SLS) et un réfractomètre. Le tampon d'élution est composé de 10mM MES à pH6,5 et 150mM NaCl.

### 3.2.1.6 Thermophorèse micro-échelle (MST)

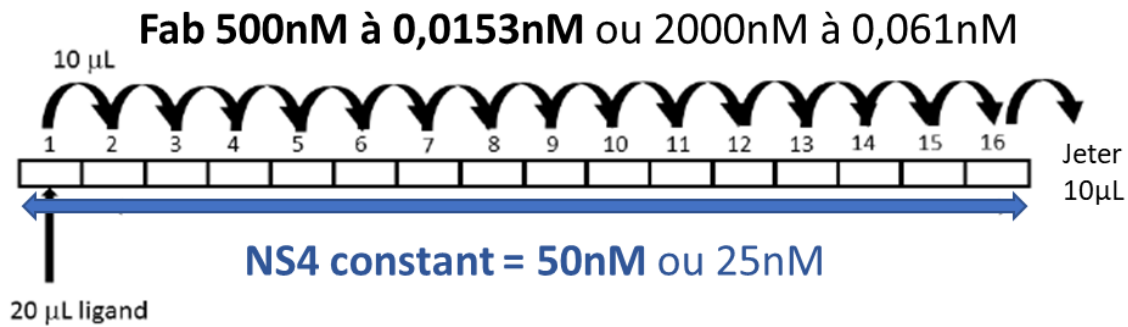
Les mesures de thermophorèse micro-échelle sont réalisées sur un Monolith NT.115 de chez Nanotemper. L'agent fluorescent RED-Tris-NTA deuxième génération est fixé sur le tag histidine de la protéine NS4. Pour cela, La protéine NS4 est diluée à 4  $\mu\text{M}$  dans un PBS+0,05 % tween. Dans une plaque 96 puits les puits 2 à 16 sont remplis avec 10  $\mu\text{l}$  de PBS-T. Dans le premier puits, 20  $\mu\text{l}$  de NS4 dilué sont mélangés avec le tampon PBS-T. Ensuite par dilutions en cascade, 10  $\mu\text{l}$  du premier puits sont distribués dans le second puits et ainsi de suite jusqu'au puits N°16. NS4 est donc diluée de 2  $\mu\text{M}$  à 0,06 nM. Le colorant RED-tris-NTA est dilué à 50 nM et 10  $\mu\text{l}$  sont ajoutés dans tous les puits. Chacun des contenants des puits est chargé dans les capillaires et la mesure est lancée à 5 % de puissance d'excitation et 20 % de puissance infrarouge le tout est réalisé à 25 °C. L'affinité mesurée est de 2 nM et 7 nM au cours de deux essais distincts (figure 3.2.1.5-1).



**Figure 3.2.1.5-1.** Mesure MST de l'affinité entre le colorant RED-tris-NTA et NS4 par MST. En (a) signal de fluorescence brut, en (b) l'analyse et extraction du Kd. Les deux expériences sont chacune associées à une couleur, exceptées pour les données en gris qui sont les mesures non considérées pour l'analyse.

A partir de ces mesures, plusieurs tests d'interactions entre le Fab-795 et NS4 sont réalisés, selon les protocoles A et B. Dans le protocole A, 90  $\mu\text{l}$  de colorant dilué à 100 nM sont mélangés avec 90  $\mu\text{l}$  de NS4 diluée à 200 nM puis ils sont incubés pendant 30 minutes à température ambiante et à l'abri de la lumière. Après quoi l'échantillon est centrifugé 10min à 4 °C à 15000g. Il en résulte donc 180  $\mu\text{l}$  de protéine marquée à 100  $\mu\text{M}$ . Le ligand (Fab-795) est dilué à 1000 nM dans le tampon de NS4. Dans une plaque 96 puits, 10  $\mu\text{l}$  de tampon de NS4 sont ajoutés dans les puits 2 à 16. Ensuite, 20  $\mu\text{l}$  de Fab-795 sont transférés dans le puits 1 avant de procéder à une dilution en cascade à partir de 10  $\mu\text{l}$  du premier puits jusqu'au puits 16. Enfin, 10  $\mu\text{l}$  de la protéine sont ajoutés dans tous les puits. La concentration finale de protéine dans les puits est de 50 nM et les Fab-795 de 500 nM à 0,0153 nM (figure 3.2.1.1.4-2).

La même procédure est réalisée avec une concentration plus élevée de Fab-795 de 2  $\mu\text{M}$  à 0,122 nM, ainsi qu'avec une composition en DDM plus faible dans le tampon. Les mesures ont été réalisées sans agent réducteur pour limiter l'impact sur le Fab.



**Figure 3.2.1.5-2.** Schéma de la dilution successive de la préparation des échantillons de NS4 et du Fab pour les analyses par MST.

### 3.2.1.7 Production in-vitro des différents variants de ORF1 et protocole des Western blots.

Afin de tester l'interaction des anticorps avec NS4, la forme polyprotéine contenant NS4 a été produite par le système d'expression cell-free. Ce test a été un double contrôle d'une part de la capacité des anticorps à reconnaître NS4 sous une autre forme et d'autre part de mesurer l'expression de la polyprotéine ORF1. Les deux variants *Den Haag* et *Sydney* ont été testés, chacun sous sa forme sauvage (WT) ou mutée (CA). Les mutants perdent la capacité de maturation par auto-clivage protéolytique et maintiennent ORF1 sous la forme de polyprotéine.

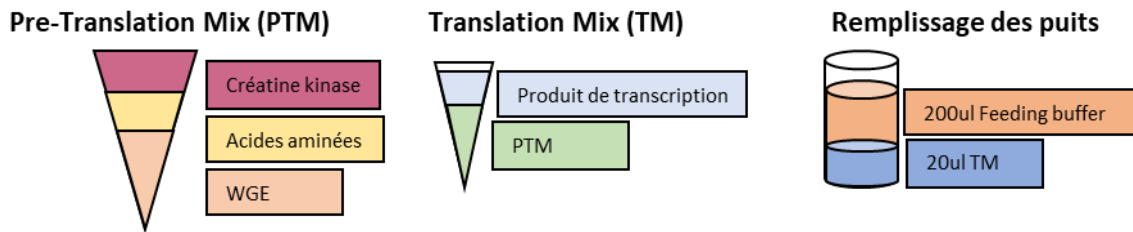
Les plasmides sont transcrits dans 4 Eppendorf RNase free pendant 7h à 37 °C dans 50 µl d'un tampon contenant 2,5mM de rNTP, 10 U/µl de SP6 polymérase, 10 U/µl de RNasine et 100ng/µl de plasmide et 10 µl de tampon de transcription.

La traduction se déroule pendant la nuit à 18 °C. La première étape consiste à créer le mélange pré-traductionnel (PTM) à une concentration de 2X contenant 80 µg/ml de créatine kinase, le mélange d'acide aminé dilué à 2X, et suffisamment d'extrait de germe de blé pour un volume total de 70 µl. Nous utilisons 11 µL de PTM pour un puits d'expression. Ensuite, le mélange traductionnel (TM) est créé en mélangeant 11 µl de PTM avec 11 µl de produit de transcription, excepté pour le témoin négatif où de l'eau RNase free est ajoutée à la place du produit de transcription.

Le tampon d'alimentation (Feeding buffer ou FB) se compose d'un mélange de quatre solutions SUBAMIX commerciales conçues pour ce système d'expression. Les quatre solutions sont mélangées puis diluées avec de l'eau distillée RNase free. Il apporte 0,3 mM d'acides aminés, 30 mM d'HEPES-KOH à pH7,6, 100 mM d'acétate de potassium, 2,7 mM d'acétate de magnésium, 16 mM de phosphate de créatine, 0,4 mM de spermidine, 1,2 mM d'ATP, 0,25 mM de GTP et 4 mM de DTT.

Enfin, chaque expression est réalisée en duplicata, ainsi les 12 premiers puits d'une plaque 96 puits sont remplis avec 200 µl de FB puis 20 µl de mélange traductionnel (TM) au fond du puits

(figure 2.1.1.5-1). La plaque est incubée sur la nuit à 18 °C.



**Figure 3.2.1.6-1.** Schéma de la préparation des différents mélanges traductionnels et du chargement des puits d'expression in-vitro.

### 3.2.1.8 Protocole du Western blot

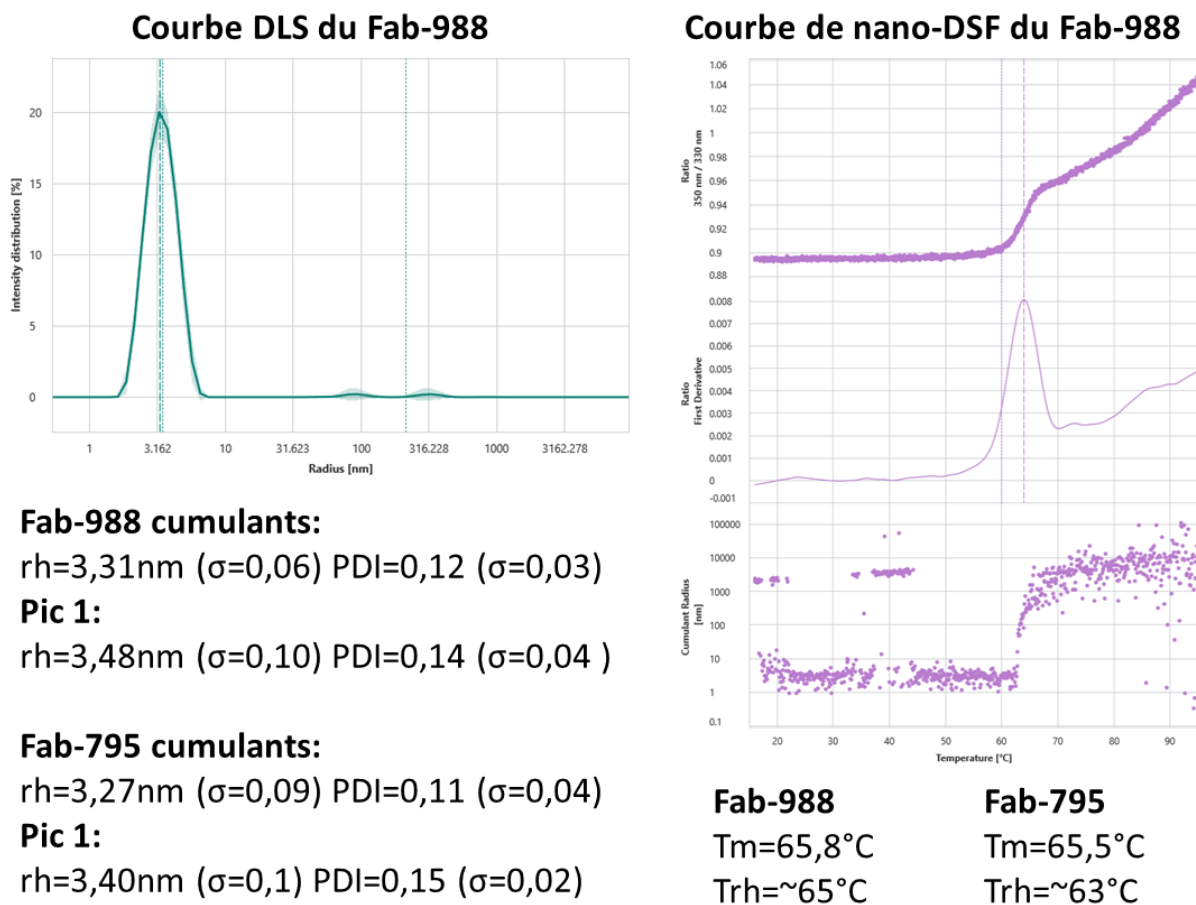
Après une nuit, les produits d'expression sont déposés en duplicata sur un gel SDS-PAGE 15 % de 15 puits en condition réductrice. Ensuite les protéines sont transférées sur une membrane de nitrocellulose durant 1h30 à 100 V. La membrane est saturée avec une solution de TBS-Tween 0,1 % et 5 % de lait pendant 1h30 sous agitation constante. La membrane est coupée en deux, une moitié est incubée 1h30 dans le mAB-795 et l'autre dans le mAB-988 dilué 30000 fois dans une solution de TBS-T et 0,5 % de lait. Les membranes sont lavées avec 50 ml de TBS-T trois fois 5 min avant de les plonger dans une solution de TBS-T à 0.5 % de lait contenant l'anticorps secondaire anti-souris couplé à la peroxydase de raifort (HRP). La HRP est une oxydoréductase qui transforme des substances chromogènes, révélant ainsi la présence de des IgG couplées à l'HRP. Les bandes sont révélées avec un luminol commercial et lu sur un appareil de lecture de luminescence.



## 3.2.2 Résultat : des conditions biochimiques incompatibles avec l'utilisation de Fab

### 3.2.2.1 Contrôle biophysique des Fabs.

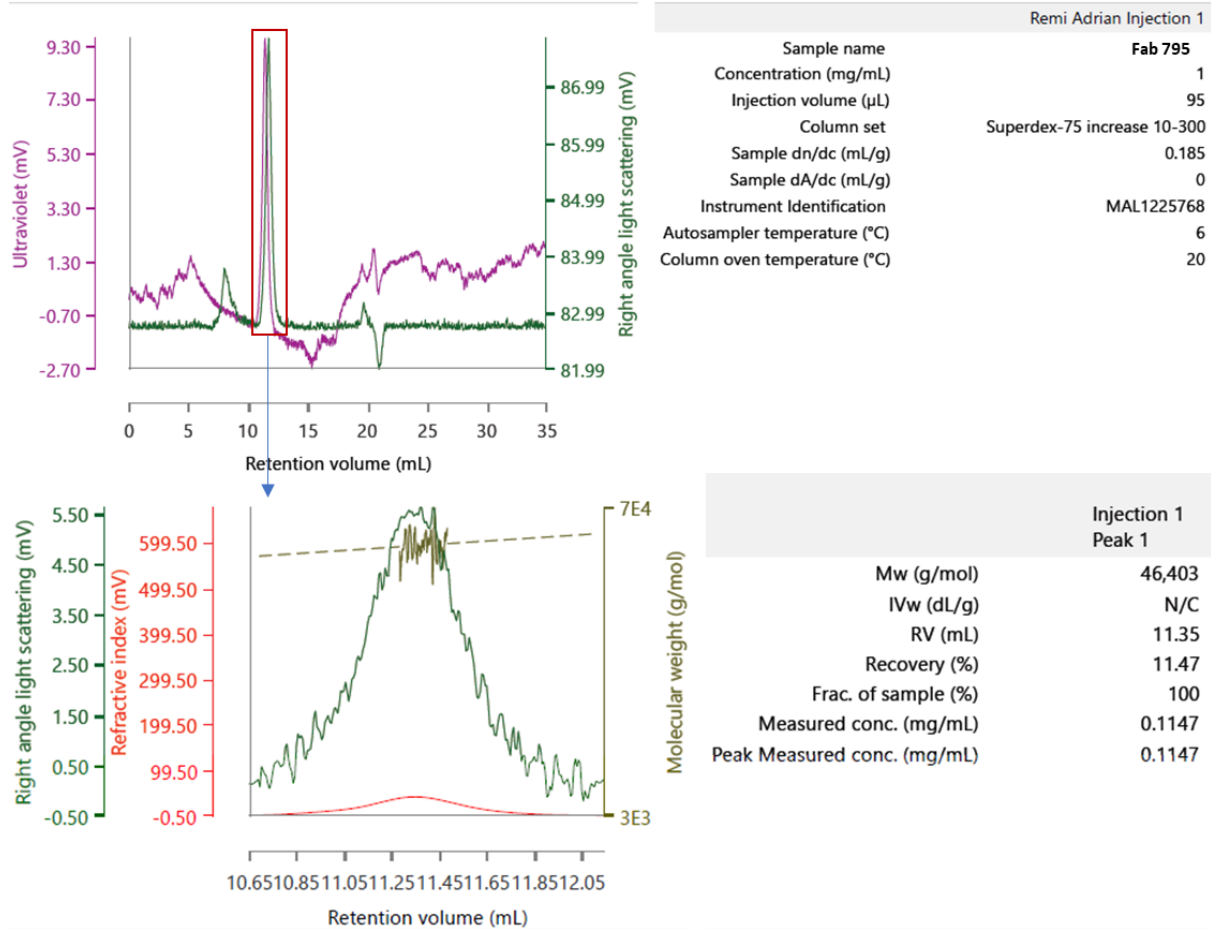
Les Fabs sont analysés par différentes techniques biophysiques comme présentés dans le chapitre précédent. En DLS les Fabs issus des deux lignées monoclonales donnent un signal de rayon hydrodynamique moyen proche de  $\sim 3,3$  nm ( $\pm 0,06-0,09$ ) et ont tous deux une température de fusion proche des  $65,6$  °C. La courbe de mesure du rayon accumulé montre une tendance à l'agrégation à partir de  $63$  °C (figure 3.2.2.1-1).



**Figure 3.2.2.1-1.** Mesure DLS et nano-DSF pour les Fabs. Les valeurs sont très proches pour les Fab-795 et Fab-988. Seules les courbes du Fab-988 sont montrées, mais les valeurs pour les deux Fabs sont indiquées sous les courbes.  $T_m$  = température de fusion,  $rh$  = rayon hydrodynamique,  $PDI$  = indice de polydispersité.

L'analyse par SEC-SLS du Fab-795, sur une colonne superdex S75 dans le tampon NS4 sans détergent (10mM MES pH6,5 + 150mM NaCl) révèle un profil d'élution particulièrement homogène et bien défini. L'analyse par réfractométrie permet d'attribuer une masse moléculaire proche des 46 kDa (figure 2.2.1-2). Enfin une mesure par MALDI-TOF permet de dévoiler un échantillon particulièrement homogène avec une masse moléculaire de 47 382

g/mol pour le Fab-795 et 47 141 g/mol pour le Fab-988. Nous pouvons noter la proximité de ses mesures avec celles de SEC-SLS (figure 3.2.2.1-3).



**Figure 3.2.1.2.1-2.** Mesure SEC-SLS du Fab dans 10mM de MES à pH 6,5 et 150mM de NaCl. Le tableau indique les mesures effectuées sur le pic encadré en rouge. Les mesures sont : le poids moléculaire (MW), le temps de rétention (RV), la part du signal associée au pic (Recovery), la part de l'échantillon dans le pic (Frac. of sample), et la concentration dans la fraction du pic (Peak measured conc.).

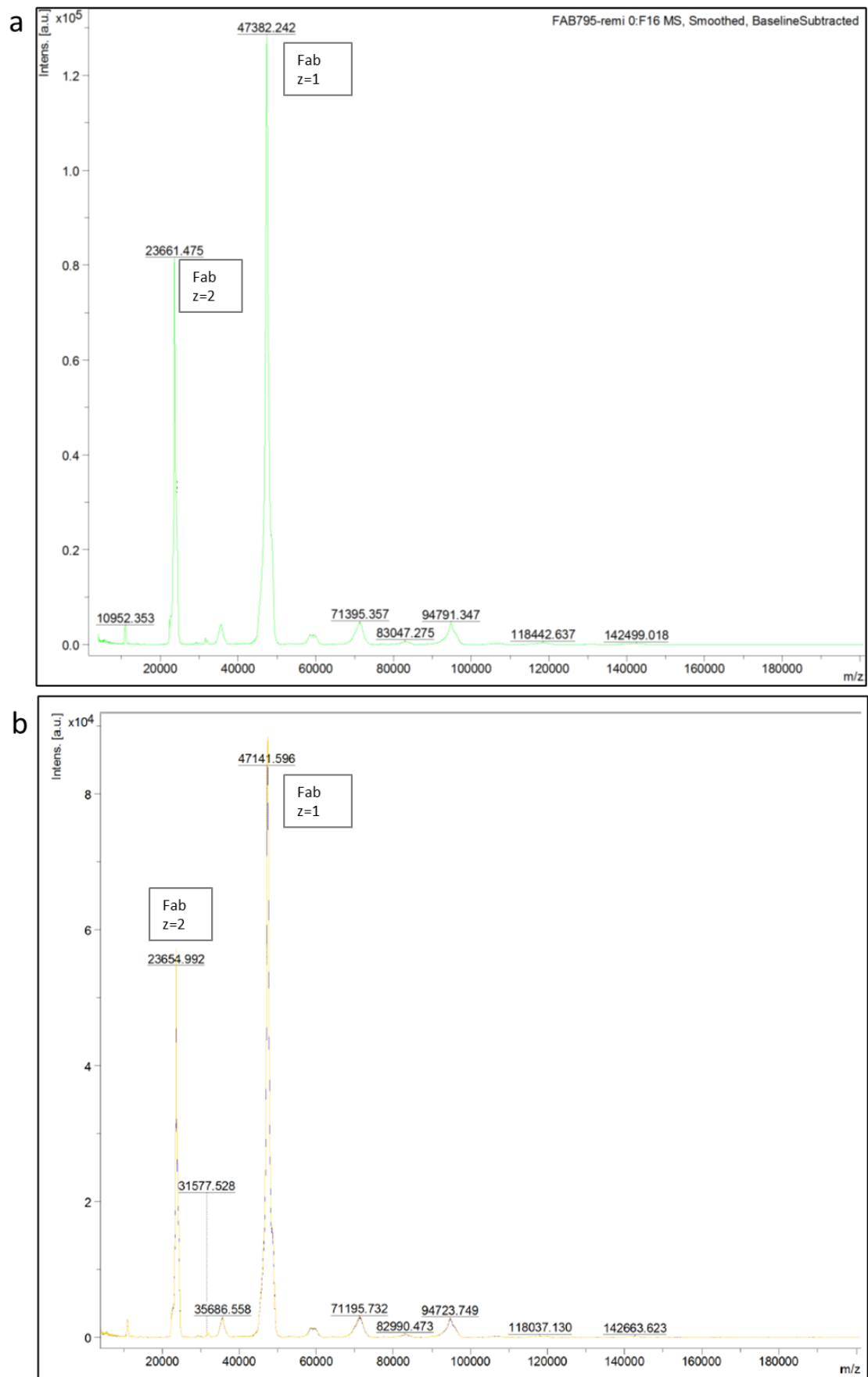
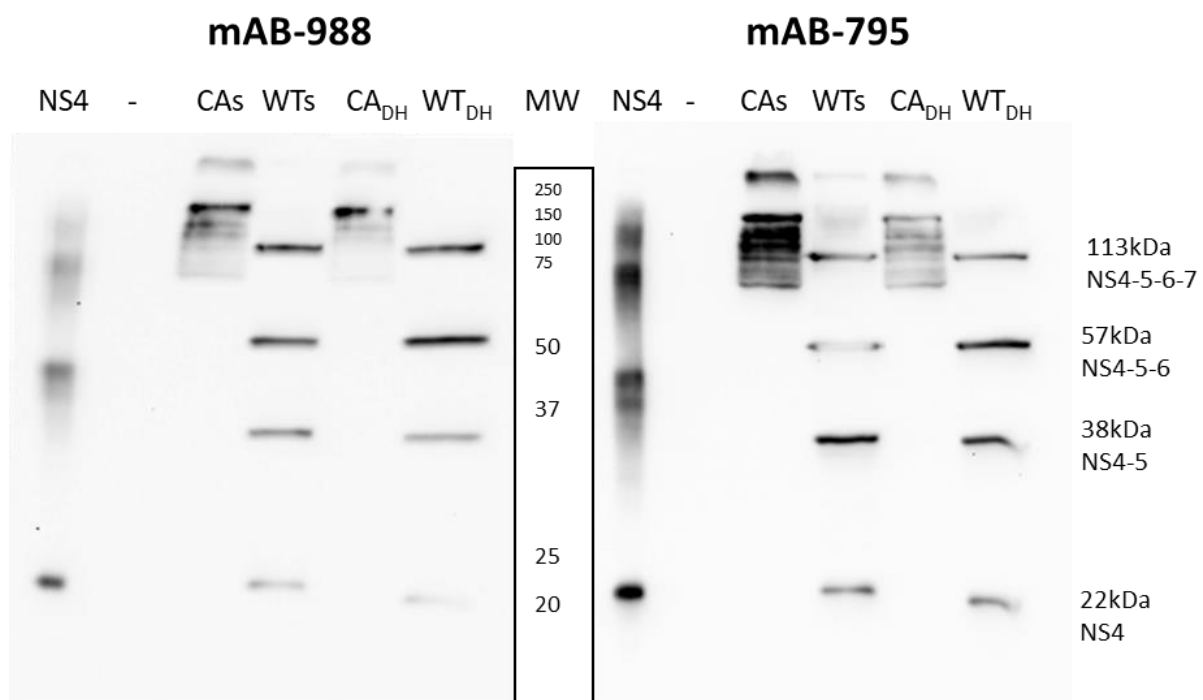


Figure 2.1.2.1-3. Résultats de l'analyse par MALDI-TOF du Fab-795 (a) et 988 (b).

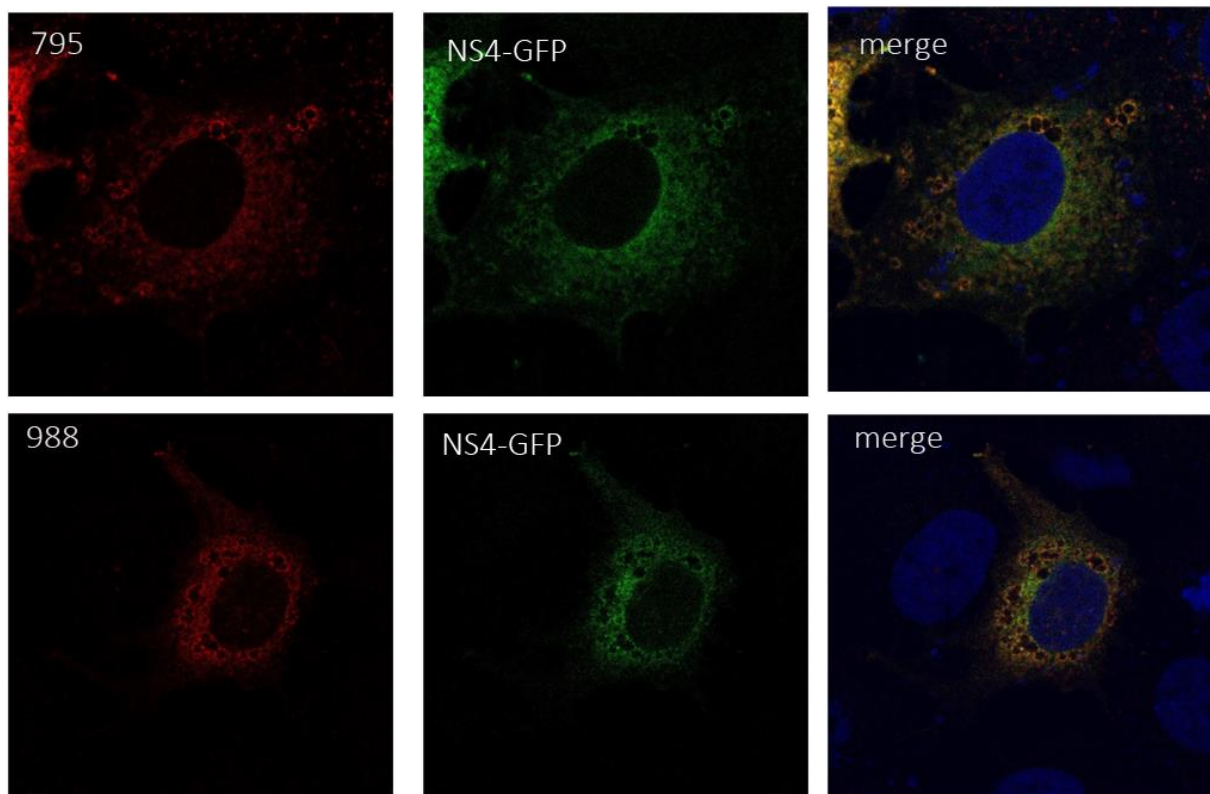
### 3.2.2.2 Tests d'interactions par Western blot et microscopie confocale.

La capacité de fixation des mABs-795 et 988 est testée en Western blot sur NS4<sub>FL</sub> classique, ainsi que sur deux variants de la polyprotéine entière ORF1<sub>Sydney</sub> et ORF1<sub>DenHaag</sub> en version sauvage (WT) ou mutée (CA). Pour ces cinq protéines nous sommes capables de révéler différentes bandes aux poids moléculaires attendus. Les différents états oligomériques de NS4<sub>FL</sub> jusqu'à au moins un état tétramérique sont visibles malgré les conditions réductrices. De même, les mutants (CA<sub>S</sub> et CA<sub>DH</sub>) ont tous une bande principale à 150kDa qui correspond à la taille attendue de la polyprotéine ORF1. Ils ont également des bandes secondaires en dessous qui peuvent être des versions partiellement traduites ou des sous-produits de dégradations. La bande au-dessus correspond à un signal d'agrégation. Concernant les versions sauvages (WT<sub>S</sub> et WT<sub>DH</sub>) nous retrouvons les différents intermédiaires de clivage portant NS4 à savoir NS4-NS5-NS6-NS7 à 113kDa, NS4-NS5-NS6 à 57kDa, NS4-NS5 à 38kDa et NS4 à 21kDa. Ces résultats peuvent être mis en parallèle avec ceux disponibles dans la littérature (Habersetzer et al. 2020). Peu d'artefacts sont visibles malgré l'utilisation du produit d'expression brut sans étape de purification indiquant une excellente spécificité (figure 3.2.2.2.-1). Ces résultats confirment avant tout la capacité des anticorps à correctement se fixer sur la protéine NS4 sous ses différentes formes monomérique, multimérique et polyprotéiques dans les conditions du Western blot. Ils ajoutent également des précisions sur la maturation de la polyprotéine ORF1 en production *in-vitro*.



**Figure 3.2.2.2-1.** Résumé des Western blots révélés par les anticorps monoclonaux 988 (gauche) et 795 (droite). Le témoin négatif (-) provient du système d'expression cell-free sans plasmide. Le témoin positif est associé à NS4 variant New Orleans. CA<sub>S</sub> et CA<sub>DH</sub> = Version mutante de ORF1 du variant Sydney et Den Haag respectivement. WT<sub>S</sub> et WT<sub>DH</sub> = Version sauvage de l'ORF1 variant Sydney et Den Haag respectivement.

Nous avons partagé ces mABs purifiés avec nos collaborateurs de l'équipe de Volker Lohmann du laboratoire de virologie moléculaire d'Heidelberg afin qu'ils les utilisent pour des expériences de microscopie confocale. Les images indiquent précisément la localisation de la protéine NS4 aux abords du réticulum endoplasmique et de l'appareil de Golgi autour du noyau (figure 3.2.2.2-2). Ces observations sont cohérentes avec les expériences passées. Nous avons longuement discuté du protocole utilisé afin de s'assurer de l'absence de faux positif. Le mAB-795 est depuis utilisé en routine dans ces expériences.



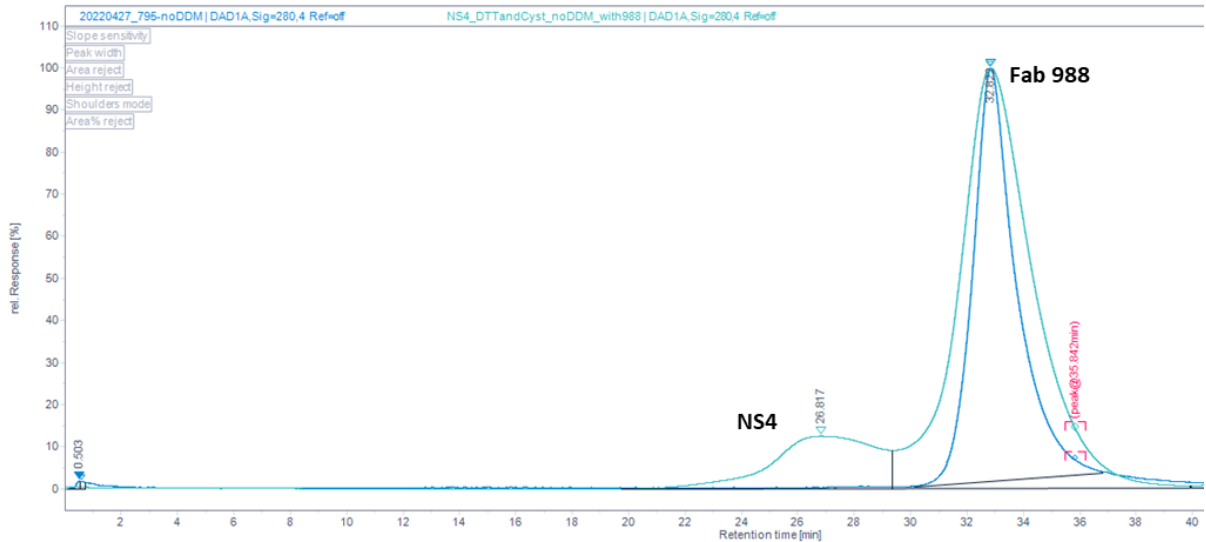
**Figure 3.2.2.2-2.** Images obtenues par Marit Lingemann du département des maladies infectieuses de l'université d'Heidelberg. Les images montrent la colocalisation de NS4 fusionnée à la GFP (milieu – vert) avec les anticorps monoclonaux 795 et 988 révélés par le fluorophore Alexa 568 via un anticorps secondaire anti-souris (gauche – rouge). Les noyaux sont colorés en bleu par le colorant Dapi. L'image présentée est une tranche unique en z sans compression. Le marquage est réalisé par perméabilisation des cellules Huh7 lunet T7 exprimant la protéine NS4.

Nous pouvons donc à ce stade affirmer que les anticorps sont capables de se fixer à NS4 dans les conditions dénaturantes du Western blot ainsi que "*in-vivo*" dans des cellules mutées. Nous allons à présent basculer vers les tests d'interaction en conditions biochimiques.

### 3.2.2.3 Test d'interaction par FPLC, AKTA et MST.

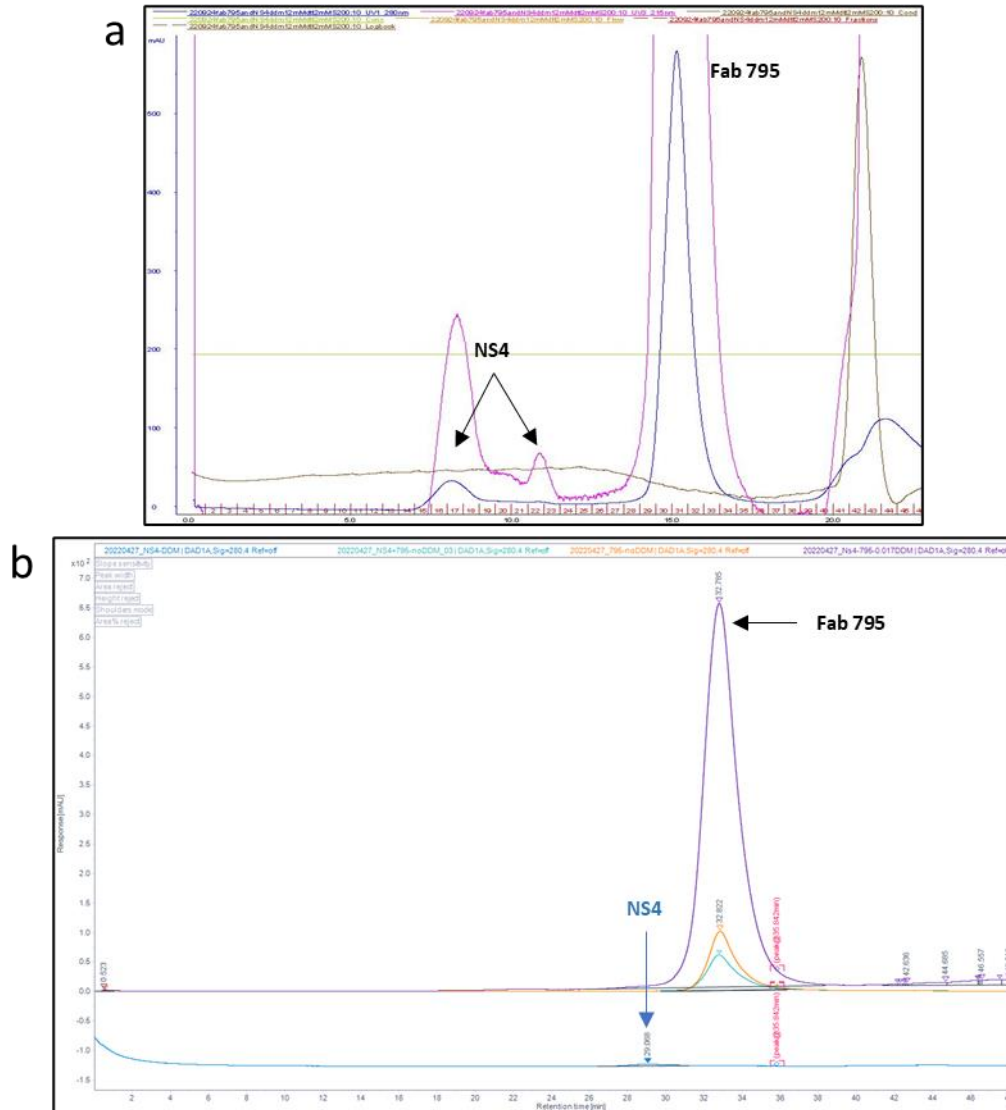
Les tests d'interaction de la protéine NS4 avec les Fabs en conditions biochimiques débutent par différentes injections en FPLC en faisant varier la concentration en détergent, les

conditions oxydantes du tampon, ainsi que le ratio Fabs/NS4<sub>FL</sub>. Dans tous les tests les Fabs sont introduits en excès dans un ratio molaire de 1NS4 pour 1,5Fab. Les deux protéines sont mélangées puis incubées environ 1 heure dans la glace avant d'être injectées. Les différentes injections dans le tampon stable identifié précédemment (0.008 % DDM et 2 mM Cystéine) ne permettent pas d'obtenir une co-élution de NS4 avec le Fab-988 (figure 3.2.2.3-1).



**Figure 3.2.2.3-1.** Profil d'élution du mélange NS4/Fab-988 dans un tampon contenant 0,008 % de DDM et 2mM de cystéine. En cyan le profil d'élution de NS4+Fab et en bleu le profil d'élution du Fab 988 seul.

Les injections à 0,006 % de DDM et 2mM de DTT et 0,017 % de DDM ne montrent pas de fixation du Fab-795 non plus (figure 3.2.2.3-2).



**Figure 3.2.1.2.3-2.** Bilan des injections NS4/Fab-795. Le profil d'élution du haut (a) montre une injection dans un tampon avec 0.006 % de DDM et 2 mM de DTT. Le tracé en bleu montre l'absorbance à 280 nm et celui en rose à 215 nm. En dessous (b) différents profils d'élution FPLC d'un mélange NS4/Fab-795 dans un tampon à 0.017 % de DDM (tracé en violet) ou sans DDM (tracé en cyan). Nous y retrouvons également le profil d'élution du Fab795 seul (orange) et NS4 dans 0.04 % DDM.

En résumé, aucune des injections n'a permis la fixation des Fabs avec la protéine (tableau 3.2.1.2.3-1). Les Fabs venant à manquer, nous avons stoppé les essais FPLC ou Akta et nous nous sommes orientés vers des expériences en thermophorèse micro-échelle.

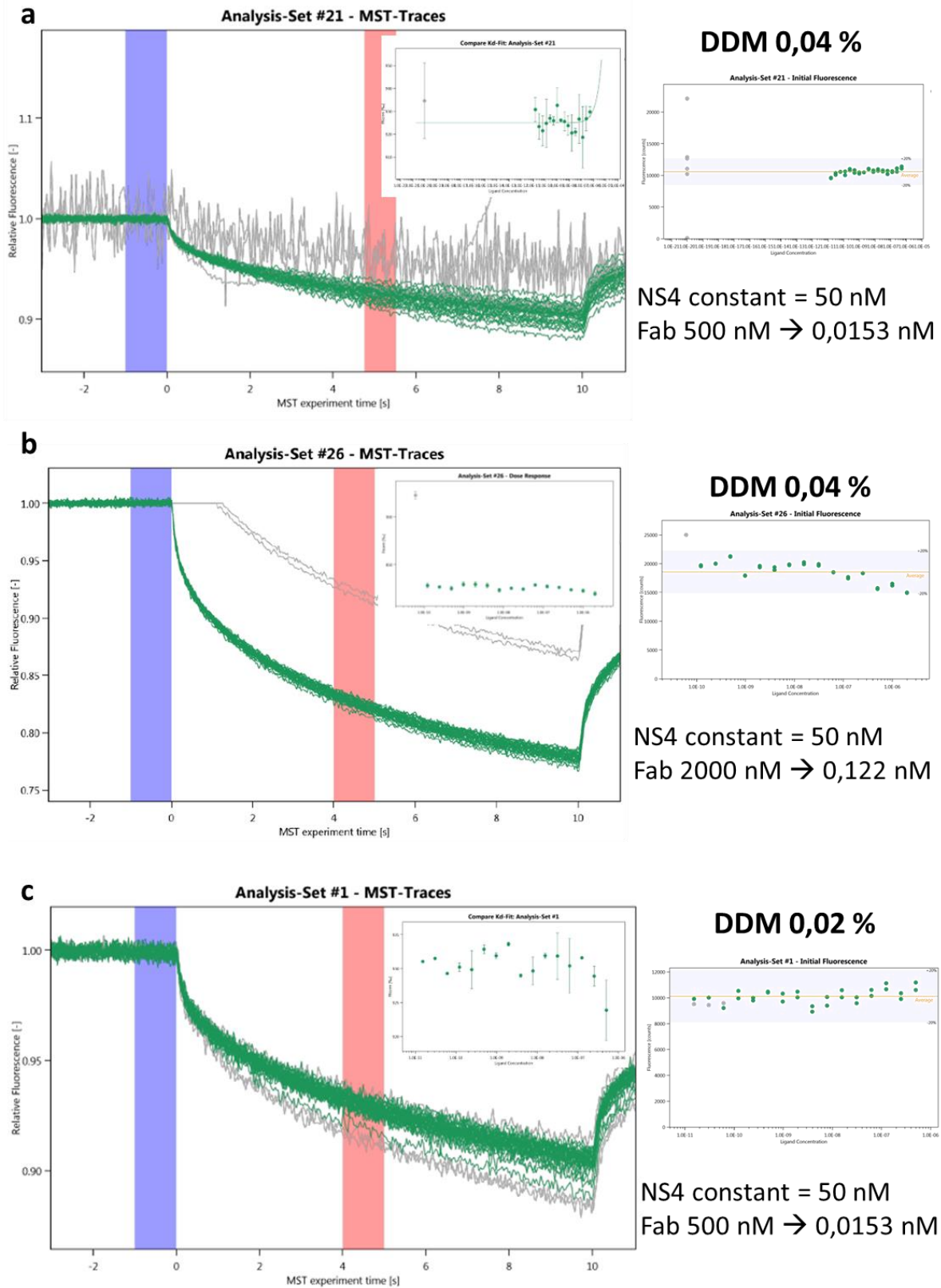
<b>Tampon</b>	<b>Quantité de protéine</b>	<b>Temps de rétention (min)</b>	<b>Masse moléculaire apparente (rayon de Stoke)</b>
Sans DDM	Fab-795 [19,5 $\mu\text{M}$ / 10 $\mu\text{l}$ ]	32,8	31 kDa (2,5)
0,017 % DDM	NS4 + 795 [40 $\mu\text{M}$ / 25 $\mu\text{l}$ ]	32,8	31 kDa (2,5)
Sans DDM	NS4 + 795 [14,7 $\mu\text{M}$ / 10 $\mu\text{l}$ ]	32,8	31 kDa (2,5)
0,008 % DDM + 2mM Cystéine	NS4 + 988 [96 $\mu\text{M}$ / 30 $\mu\text{l}$ ]	26,8 + 32,8	124 kDa + 31 kDa
0,008 % DDM + 2mM Cystéine	NS4 [96 $\mu\text{M}$ / 30 $\mu\text{l}$ ]	26,1	146kDa

**Tableau 3.2.2.3-1.** Récapitulatif des injections sur FPLC.

Pour les expériences en MST, la protéine NS4 est préalablement marquée avec un fluorophore via le tag histidine comme décrit dans le matériel et méthode. Ensuite la protéine et le Fab-795 sont incubés pendant 10min avant de lancer l'expérience. Le premier test est effectué dans un tampon contenant 0,04 % de DDM, NS4 est à une concentration constante de 50 nM et le Fab-795 entre 500 nM et 0,0153 nM. Cette gamme est choisie en fonction de l'affinité attendue pour un anticorps qui doit être de l'ordre du nano-molaire. La fluorescence varie de façon identique pour chaque condition indiquant qu'il n'y a pas eu d'interaction. L'expérience est donc reproduite avec cette fois-ci une gamme de Fab allant de 2  $\mu\text{M}$  à 0,122 nM dans le même tampon mais conduit aux mêmes observations (figure 3.2.2.3-3). Le contrôle de la fluorescence initiale permet d'écartier la formation d'agrégats tout en garantissant le marquage correct de la protéine NS4.

Le dernier test est réalisé en abaissant la concentration du détergent à 0,02 % avec une concentration de Fab de 500 nM à 0,0153 nM correspondant à deux fois la CMC. Encore une fois aucune fixation des Fab n'est repérée dans cette condition (figure 3.2.2.3-4).





**Figure 3.2.2.3-4.** Résultats des mesures MST du mélange NS4/Fab-795. En **a** les données brutes en condition 0,04 % de DDM, en **b** les données avec le Fab-795 concentré à 2  $\mu$ M et 0,04 % de DDM et enfin en **c** les données du Fab-795 mesuré avec 0,02 % de DDM.

### 3.2.3 Conclusion et discussion : les dérivés d'anticorps ne sont pas des outils sans failles pour l'étude structurale

Nous avons vu par une analyse biophysique poussée des Fabs que ceux-ci sont d'excellente qualité en tous points. Il est intéressant de noter la cohérence des mesures les unes avec les autres. Pour chaque Fab les mesures par DLS, SLS et MALDI-TOF convergent vers un échantillon homogène d'une taille d'environ 47kDa.

Nous avons également mis en opposition les résultats obtenus par Western blot et en microscopie confocale avec ceux obtenus en biochimie. Il y a deux différences majeures entre les moments où l'interaction est visible et ceux où elle ne l'est pas. La première distinction est que pour le Western blot et les observations en cellules, ce sont les anticorps entiers qui sont utilisés. Alors que, les tests biochimiques sont réalisés avec les Fabs. La deuxième distinction est l'environnement de la protéine NS4. Dans les Western blot la protéine est dans un état dénaturé. En cellule elle est dans un environnement biologique, associée à des endomembranes du côté du cytoplasme cellulaire. Tandis que dans les conditions biochimiques, NS4 est en solution en présence de détergent DDM qui forme de nombreuses micelles avec lesquelles la protéine interagit. Nous espérons que les Fab puissent masquer la région hydrophobe de NS4 et aider à sa solubilisation en agissant comme un détergent mais ce n'est pas le cas. Nous avons vu lors des mesures SEC-SLS sur NS4 que la micelle de détergent prend une part importante du signal elle pourrait donc induire une gêne stérique qui empêche la fixation du Fab. Dans ce cas, le problème réside dans la sélection des anticorps monoclonaux qui est faite dans des conditions in-vivo et donc éloignées de la biochimie. Il est aussi possible que la transition de l'anticorps vers le Fab réduise drastiquement l'avidité sur les formes oligomériques de NS4 du fait que le Fab ne présente qu'un seul Paratope.

Nous avons également illustré dans ce chapitre les limites typiquement rencontrées lors de l'utilisation de mABs. Les mABs sont chers à produire et nécessitent un traitement pour produire les Fabs qui est de faible rendement (30 à 40 %). Nous disposons donc d'une faible quantité de Fabs qui sont rapidement venus à manquer et auraient rapidement pu faire exploser le coût des expériences. C'est ce qui est à l'origine du nombre limité de tests réalisés. C'est pour cette raison que les mABs ne sont pas l'outil idéal pour une application purement à but d'analyse structurale.

Pour pallier ce manque, nous avons décidé d'obtenir un catalogue diversifié d'alpharep qui ont déjà trouvé une application en cristallographie et qui seront également utiles en cryo-EM. Leur mode de sélection par phage display se fait en condition biochimique, et nous permet de résoudre le décalage supposé entre la sélection et l'application.



## 3.3 Les Alphareps dirigées contre la protéine NS4

### 3.3.1 Matériel et méthode

#### 3.3.1.1 Sélections des alphareps par la technique du phage display

Les alphareps sont sélectionnées par phage display. C'est la séquence NS4<sub>FL</sub> du génogroupe II génotype 4 variant Helensbourg qui est utilisée pour la sélection. La banque de phage du départ a été développée préalablement par l'équipe de modélisation et ingénierie des protéines de l'I2BC (Urvoas et al. 2010; Urvoas et al. 2012; Valerio-Lepiniec et al. 2015). La banque naïve, qui correspond à la diversité en alpharep au départ de la sélection, est estimée à environ  $1,7 \cdot 10^9$  clones indépendants.

##### 3.3.1.1.1 Protocole de sélection des alphareps

La sélection est divisée en trois tours qui s'étalent sur trois à quatre jours chacun.

#### Jour 1.

**Préparation des milieux et boîtes de culture.** Le milieu de culture est le 2x yeast extract-tryptone (2YT) spécialement conçu pour la culture de phage M13. Les boîtes sont composées de 2YT, ampicilline, tétracycline, du glucose à 1 % et de l'agar. La résistance à la tétracycline est portée par les bactéries XL1 blue. La résistance à l'ampicilline est apportée par les phages M13. Les phages M13 sont des phages non lytiques contenant le phagemid avec la séquence de l'alpharep. Le glucose permet d'inhiber le promoteur T7 et limiter les fuites d'expression des alphareps pendant certaines étapes. Toutes ces préparations sont stérilisées via un chauffage raisonné (90 °C) réalisé manuellement dans une cocotte-minute.

**Début de la dialyse du stock de phage naïf.** Les phages issus de la banque « naïve » sont dialysés dans le tampon de la protéine sans DDM ni DTT (20mM MES pH 6,5 et 150mM NaCl). Les phages sont dialysés dans 1L pendant une journée puis dans un nouveau litre sur la nuit le tout à 4 °C. Une fraction des phages non dialysés est récupérée pour être archivée.

**Préparation des boîtes 96-puits non spécifique (NS) et spécifique (S).** Les deux boîtes sont toutes revêtues de streptavidine par dépôt de 100 µl d'une solution à 20µg/ml dans du PBS préalablement préparé. La plaque est ensuite centrifugée pendant 3heures à 300rpm à 20 °C. Chaque puits est lavé 3 fois avec 300 µl de TBS-Tween 0,1 % w/v (0,1g de tween pour 100 ml). Enfin, la surface est saturée avec 300 µl/puits de TBS-T0,1 % + BSA4 % et incubée sur la nuit à 4 °C sous agitation à 300rpm.

**Lancement de la préculture des XL1 blue MRF'.** Une préculture de XL1 blue MRF' dans 5 ml de 2YT-Tetracycline (qui servira pour le clonage) est incubée à 37 °C sous agitation.

#### Jour 2

**Fin de la dialyse.** Le volume de tampon contenant les phages dialysés sur la nuit est complété avec du DDM jusqu'à 0,02 % et avec du DTT jusqu'à 2mM avant d'être répartis en Eppendorf

dans des fractions de 1 ml.

**Culture XL1 blue pour titration.** On inocule 20 ml de 2YT-T avec la préculture de XL1blue pour atteindre une densité optique d'environ 0,15. Une fois que la culture a atteint une  $DO_{600}$  de  $\sim 0,8$ , elle est utilisée pour la titration.

**Montage du NTA-Biotin sur la plaque spécifique.** Une solution de NTA-Biotin à 1  $\mu\text{g/ml}$ , 5 mM de  $\text{NiSO}_4$ , dans du TBS est déposée dans les puits de la plaque S et incubée 30min à 20 °C et 300rpm. La plaque spécifique est lavée avec 3x300  $\mu\text{l}$ /puits de tampon tris salin complémenté avec 0,1 % de tween (TBS-T0,1 %). Après le lavage la plaque est incubée avec 100  $\mu\text{l}$ /puits de NiNTA-Biotin pendant 45min à 20 °C et à 300rpm.

**Sélection des phages non-spécifiques.** La plaque NS est lavée avec 3x300  $\mu\text{l}$ /puits avec le tampon spécifique (20mM MES, 150mM NaCl 0,02 % de DDM et 2mM DTT). Puis 130  $\mu\text{l}$ /puits de phages provenant de la dialyse dans le même tampon sont ajoutés en utilisant des cônes à filtre, avant d'incuber pendant 2h à 20 °C et 300rpm.

**Fixation de la cible (NS4) sur la plaque spécifique.** Les puits de la plaque spécifique sont enduits par 100  $\mu\text{l}$  de NS4 à 30 $\mu\text{g/ml}$  puis la plaque est incubée pendant 2h à 4 °C et 300rpm.

**Fixation des phages.** La plaque spécifique (contenant NS4) est lavée avec 3x300  $\mu\text{l}$ /puits de tampon spécifique (20mM MES pH6,5, 150mM NaCl, 0,02 %DTT et 2mM DTT). Ensuite, 100  $\mu\text{l}$  de phages dans le même tampon sont transférés depuis la plaque NS vers la plaque S en utilisant des cônes avec filtre. La plaque est incubée 2h à 4 °C et 300rpm.

**Lavage et élution des phages.** Après incubation la plaque spécifique est lavée avec 20x300  $\mu\text{l}$  de tampon spécifique. Le 21<sup>ème</sup> tour de lavage est fait avec 100  $\mu\text{l}$  de tampon spécifique et est récupéré pour la titration. L'élution est réalisée en ajoutant 100  $\mu\text{l}$  de glycine à 100mM et pH 2,5, puis en incubant pendant 10min à 20 °C et 300rpm. L'élution est finalement récupérée et l'acidité est neutralisée avec 80  $\mu\text{l}$  de tampon tris 1M à pH 7.

**Infection des XL1Blue MRF' et titration.** Les phages issus de l'élution sont dilués en cascade avec 10  $\mu\text{l}$  dans 990  $\mu\text{l}$  pour les sauts par  $10^{-2}$  et 100  $\mu\text{l}$  dans 900  $\mu\text{l}$  pour les sauts  $10^{-1}$ . Les différentes dilutions sont résumées ci-après :

Titration des phages entrant dans le tour 1 (de la dialyse) INPUT : Dilutions  $10^{-7}$ ,  $10^{-8}$  et  $10^{-9}$

Phages du lavage L21 : Dilutions  $10^{-2}$ ,  $10^{-4}$ ,  $10^{-6}$

Phages de l'élution OUT : Dilutions  $10^{-2}$ ,  $10^{-4}$ ,  $10^{-6}$

Les bactéries sont infectées avec 100  $\mu\text{l}$  d'un lot de phages dilués pour infecter 200  $\mu\text{l}$  de bactéries à  $DO_{600}=1$  issues de la culture. D'un autre coté 500  $\mu\text{l}$  de phages élués sont ajoutés

à 1,5 ml de bactéries puis sont incubés 30min à 37 °C sous agitation. Les bactéries infectées sont ensuite étalées dans les boîtes de Petri 2YT-AT-Glucose1 %. Dans les petites boîtes seuls 200 µl de bactéries infectées sont utilisées et l'intégralité des 2 ml de bactéries éluées sont étalées dans une grande boîte de Petri de dimension 12cmx12cm.

### **Jour 3**

**Réplication des phages sélectionnés via les phages helper K1.** Les clones sont repris de la grande boîte dans 7 ml+3 ml de 2YT-Glucose 1 %-Glycérol 15 % pour l'ensemencement de 50 ml de 2YT-AT-Glucose1 % à une DO d'environ 0,15. Le reste des bactéries est stocké à -80 °C. Afin de débiter la phase d'infection, une fois la culture à ~0,7-0,9, nous ajoutons 1,8 ml de phage helper K1 concentré à environ  $10^{11}$  phages/ ml. Ces phages apportent la capacité de réplication aux phages ainsi que la résistance à la kanamycine. Ainsi le milieu de culture pourra sélectionner les bactéries infectées par les phages helper. Ce mélange est laissé au repos pendant 30min à 37 °C puis sous agitation pendant 30min à 37 °C et 100rpm pour permettre une infection optimale. Afin de permettre la production de phage la solution de bactéries infectées est centrifugée pendant 15min à 20 °C et 4000 rpm. Le culot est repris dans 50 ml de 2YT-AK puis incubé sur la nuit à 37 °C sous agitation constante à 100rpm.

**Comptage des clones des différentes dilutions.** Pour chaque dilution étalée sur boîte la veille le comptage des clones permet le suivi de la sélection de phage. Les comptes de chaque tour sont résumés dans le tableau 3.1.1-1. Les phages issus du lavage, de l'éluion et de la grande plaque sont congelés et stockés dans le -80 °C jusqu'à la fin du processus de sélection.

Les tours 2 et 3 sont une répétition du tour 1 sauf quelques exceptions. Le stock de phage de départ correspond aux phages obtenus après éluion en fin de sélection du tour précédent. Pour le dernier tour de sélection le lavage est fait 41 fois au lieu de 21 fois.

	Tour 1			Tour 2			Tour 3		
	10 <sup>-7</sup>	10 <sup>-8</sup>	10 <sup>-9</sup>	10 <sup>-6</sup>	10 <sup>-7</sup>	10 <sup>-8</sup>	10 <sup>-7</sup>	10 <sup>-8</sup>	10 <sup>-9</sup>
INPUT	100	6	0			ND 220 / D <1000			73
L21/41	10 <sup>-0</sup>	10 <sup>-1</sup>	10 <sup>-2</sup>	10 <sup>-2</sup>	10 <sup>-3</sup>	10 <sup>-4</sup>	10 <sup>-2</sup>	10 <sup>-3</sup>	10 <sup>-4</sup>
	18	0	0		>110	18		>200	11
OUTPUT	10 <sup>-0</sup>	10 <sup>-1</sup>	10 <sup>-2</sup>	10 <sup>-2</sup>	10 <sup>-3</sup>	10 <sup>-4</sup>	10 <sup>-2</sup>	10 <sup>-3</sup>	10 <sup>-4</sup>
	40	17	0		>400	57		>400	46
Régénération	~1500 clones			Tapis = DO 70			Tapis DO=35.8		

**Tableau 3.1.1-1.** Résumé des comptages des colonies en fonction des dilutions pour chaque tour. INPUT= Phage après dialyse, L21/41= Phages au dernier cycle de lavage, OUTPUT= Phages après élution, Régénération= Phages obtenus sur la grande plaque et la densité optique de la culture résultante. Pour la colonne du tour 2 ND=Non dialysé et D=Dialysé.

### 3.3.1.1.2 Préparation de la plaque matrice et mise en culture des clones sélectionnés

Une fois les phages sélectionnés les clones positifs sont révélés par test ELISA.

**Ensemencement de la plaque matrice.** Dans une plaque 96 puits, nous ajoutons 150 µL de 2YT-AT-Glucose 1 % par puits. Chaque puits est ensemencé par un clone différent. Les deux premières lignes (A et B) sont ensemencées par des colonies du tour 2 et les autres lignes avec des colonies du tour 3. Les bactéries sont mises en croissance la nuit à 37 °C à 300rpm. 30 µl de cette préculture sont ensemencés dans 140 µl de 2YT-AT-Glucose puis incubés pendant 4h à 37 °C et 600rpm.

**Infection par les phages helper K1.** Pour rappel les phages helper apportent la capacité de réplication au phage ainsi que la résistance à la kanamycine. Ils sont dilués deux fois dans du 2YT et 10 µl/puits sont ajoutés en utilisant la pipette multicanaux. La plaque est ensuite incubée 30min à 37 °C sans agitation puis 30min à 37 °C et 300 rpm afin d'optimiser l'infection des bactéries.

**Préparation de la plaque deep-well.** Dans une plaque à 96 puits profonds et stériles, nous ajoutons 1,5 ml de 2YT-Ampiciline-Kanamycine par puits avant de transférer 100 µl de la plaque matrice préalablement préparée vers la plaque deep-well. La plaque est scellée avec un film stérile et mise en culture sur la nuit à 30 °C et 300rpm.

#### 3.3.1.1.3 Révélation par phage ELISA

Les différentes colonies vont être réparties sur deux plaques 96 puits en double. Une ligne sera préparée selon le protocole d'une surface spécifique tandis que la ligne en dessous correspondra au contrôle négatif et sera préparée comme une surface non spécifique.

**Fixation de la Streptavidine et saturation par la BSA.** De la même manière que lors des tours de sélection deux plaques 96 puits sont activées avec de la streptavidine puis saturées avec de la BSA. La plaque est d'abord incubée avec 100 µl de streptavidine à 20µg/ ml dans du TBS pendant 3h à 20 °C et 300rpm. La saturation est faite en ajoutant 300 µl/puits de BSA 4 % dans du TBS-T 0.1 % sur la nuit à 4 °C et 300rpm.

**Application du tris-NTA-NiSO<sub>4</sub>.** Le lendemain, à la manière du protocole présenté pour les tours de sélection le Ni-NTA-NiSO<sub>4</sub> est ajouté sur les lignes spécifiques uniquement et du TBS-T est ajouté sur les lignes non-spécifiques.

**Ajout de la protéine NS4.** Chaque puits est lavé trois fois avec 300 µl de tampon spécifique à NS4. NS4 est diluée pour une concentration finale de 35µg/ ml dont 110 µl sont distribués dans les puits spécifiques uniquement. La plaque est incubée 2h à 4 °C et 300rpm.

**Ajout des Phages sélectionnés.** Les plaques ELISA sont lavées avec 3x300 µl de tampon spécifique à NS4 fraîchement préparé. 10 µl d'un tampon NS4 à 11mM de DTT sont ajoutés dans les puits des deux plaques ELISA avant d'ajouter 100 µl/puits de phage depuis la plaque deep-well vers tous les puits de la plaque ELISA. Les deux plaques sont incubées 1h30 à 20 °C 300rpm. Durant cette étape les phages M13 qui expriment les alphareps vont se fixer sur NS4 plus ou moins bien selon l'affinité pour NS4 de l'alpharep exprimée.

**Ajout des anticorps anti-M13.** Les plaques sont lavées trois fois avec 300 µl de tampon spécifique à NS4 fraîchement préparé. Les anticorps anti-M13 couplé HRP sont ajoutés dans tous les puits, spécifiques et non spécifiques. Les anticorps sont dilués au 2000<sup>ième</sup> dans le tampon spécifique à NS4 (10 µl dans 20 ml) et 100 µl/puits sont ajoutés. La plaque est incubée 1h à 20 °C et 300rpm.

**Révélation de l'ELISA.** La plaque est lavée trois fois avec 300 µl de tampon spécifique à NS4 fraîchement préparé. Puis la révélation est faite à partir de substrat POD soluble par ajout de 100 µl/puits sans changer les cônes mais sans tremper dans les puits. La réaction est stoppée avec 100 µl/puits de tampon HCl à 1M.

**Séquençage des phagemides.** Parmi les puits positifs en ELISA, 50 clones sont sélectionnés pour être séquencés par le service PlateSeq de chez Eurofin (cf tableau 3.2-1). Dans une plaque spéciale multiséquençage 200 µl de milieu sont inoculés avec 10 µl de cellules issues de la plaque matrice. La plaque est ensuite incubée sur la nuit à 37 °C et 300rpm. Après la culture, la plaque est scellée et envoyée en séquençage.



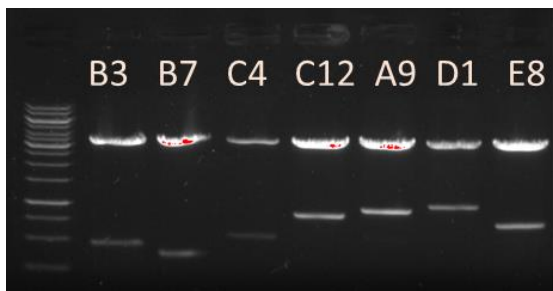
#### 3.3.1.1.4 Clonage des alphareps dans le vecteur d'expression

**Extraction des gènes depuis les phagemides.** Sept clones différents sont sélectionnés pour le clonage et l'expression. Les phagemides sont extraits à partir de cultures de 5 mlensemencées à partir de 10 µl de bactéries de la plaque matrice, mises en croissance à 37 °C sur la nuit. Le lendemain les phagemides sont extraits par le protocole de miniprep standard via le kit de chez Macherey-Nagel. Une fois purifiés, les différents produits sont dosés par une mesure UV à 260 nm (tableau 3.1.1-3)

Name	Concentration	260/280	260/230
B3	110,6 ng/µl	1,92	2,40
A9	150 ng/µl	1,94	2,49
E8	149 ng/µl	1,93	2,46
B7	113 ng/µl	1,95	2,39
C4	109 ng/µl	1,79	2,21
C12	155,6 ng/µl	1,96	2,44
D1	192,8 ng/µl	1,84	2,32

**Tableau 3.3.1.1-3.** Bilan de l'extraction des phagémides pour chacun des clones sélectionnés.

**Extraction et purification des inserts.** Les phagemides purifiés sont digérés pour extraire le gène des alphareps. Pour cela 10 µl de plasmide sont mélangés avec 5 µl de tampon de digestion commerciale, 1 µl d'enzyme HindIII, 1 µl d'enzyme BamHI, et 13 µl d'H<sub>2</sub>O. Le tout est incubé à 37 °C pendant 1h. le produit de digestion est déposé sur un gel d'agarose à 1 % (Figure 3.1.1-1). Les bandes sont découpées est l'ADN est extrait en utilisant un kit d'extraction et de purification d'ADN de chez Macherey-Nagel. Et les gènes sont dosés par analyse UV (tableau 3.1.1-4). Une fois purifiés, les gènes sont introduits dans le plasmide PQE31 par digestion du plasmide avec les enzymes BamHI et HindIII, puis en mélangeant les inserts avec le plasmide digéré dans un ratio de 37ng/50ng/1 µl insert/plasmide/Ligase puis en incubant 2h à 27 °C.



**Figure 3.3.1.1-1.** Gel agarose de la digestion des phagemides. Les bandes du haut correspondent au phagemide, les bandes du bas correspondent

aux inserts c'est-à-dire aux gènes des différentes alphareps.

Name	Concentration	260/280	260/230
B3	11,8 ng/ul	1,76	0,78
A9	25,1 ng/ul	1,72	1,23
E8	18,5 ng/μl	1,84	0,85
B7	10,6 ng/μl	1,71	0,75
C4	8,9 ng/μl	1,57	0,22
C12	21,7 ng/μl	1,78	1,03
D1	13 ng/μl	1,62	1,06

**Tableau 3.3.1.1-4.** Mesures UV des inserts après extraction et purification du gel d'agarose.

**Transformation des DH5α pour séquençage.** Trois plasmides sont utilisés (A9, C4 et B3) pour transformer les DH5α afin d'amplifier et séquencer les plasmides, seulement trois ont été séquencés. Pour cela les cellules sont incubées 10min dans la glace. Puis 2 μl de plasmide sont mélangés avec les bactéries. Le mélange est incubé dans la glace pendant 30min avant un choc thermique à 42 °C pendant 2min15. Il est à nouveau plongé dans la glace pendant 5min. 900 μl de LB stérile sont ajouté au mélange, qui est ensuite incubé à 37 °C, 200rpm pendant 80min. L'Eppendorf est centrifugé 10min à 2500rpm, puis 800 μl de surnageant sont retirés. Le culot est repris dans les 100 μl restant puis étalé sur une boîte de pétrie LB Agar ampicilline (0.1mg/ml). La boîte est incubée à 37 °C pendant la nuit.

**Purification et séquençage des plasmides.** Trois clones isolés pour chacun des trois plasmides sélectionnées (9 échantillons) sont utilisés pour ensemencer 5 ml de LB avec de l'ampicilline qui est mis en culture à 37 °C sur la nuit. Le lendemain, les plasmides sont purifiés à l'aide d'un kit miniprep de chez Macherey-Nagel. Les plasmides ainsi purifiés sont envoyés à séquencer.

**Transformation pour expression.** À partir des plasmides ayant la bonne séquence, des bactéries commerciales BL21-Gold (DE3) sont transformées selon le même protocole que celui présenté précédemment (Transformation des DH5α) en ajustant le temps d'incubation du choc thermique à 45sec et en utilisant le milieu de croissance commercial.

### 3.3.1.2 Production et purification des alphareps

**Culture des bactéries.** L'expression est réalisée à partir d'une préculture de 25 ml de LB avec de l'ampicilline, mise en croissance sur la nuit. Cette préculture sert à ensemencer 1 litre de culture dans le même milieu à une DO = 0,15 qui est mis en culture à 37 °C. Une fois que la DO atteint 0,8, la production est induite avec 1mM d'IPTG et la culture incubée pendant 4h à 37 °C et 220rpm. Les cultures sont centrifugées pendant 15 minutes à 20 °C à 5000g. Les culots sont

congelés à -20 °C.

**Purification des alphareps.** Les cellules sont lysées par sonication (4\*30sec 60 %, 1sON/1sOFF) dans le tampon de lyse contenant de la benzonase, 50 mM de NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 20mM Imidazole à pH8. Le produit de lyse est centrifugé à 10000g à 4 °C pendant 1h, le surnageant est filtré à 0,22 µM. Sur une potence, 4 ml de résine Ni-NTA sont équilibrés dans le tampon de lyse. La fraction soluble et filtrée est passée sur la colonne puis la colonne est lavée avec le tampon de lyse jusqu'à disparition du signal lors des mesures UV à 280nm. L'élution est faite dans 50mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl et 300 mM Imidazole à pH8 réparties en plusieurs fractions de 2 ml. Les fractions d'élutions sont concentrées et passées en chromatographie d'exclusion de taille sur une colonne (Superdex 75 (GE Healthcare) équilibrée dans un tampon de 50mM de phosphate de sodium, 150 mM NaCl à pH7,4.

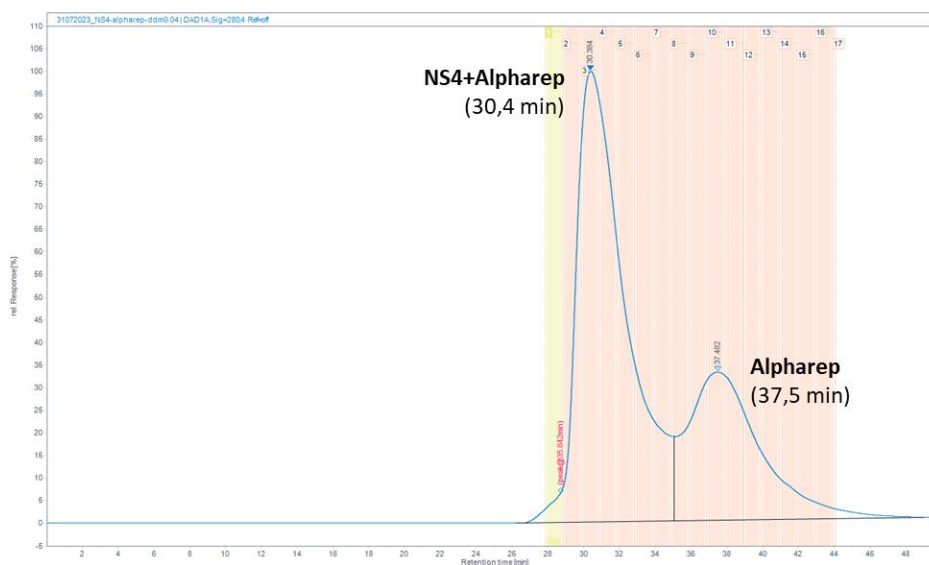
#### *3.3.1.3 Caractérisation biophysique des alphareps*

Les analyses par nano-DSF et DLS sont effectuées de la même manière que décrit dans les chapitres précédents. Les mesures sont réalisées sur l'appareil Prometheus Panta. Les mesures de nano-DSF sont faites sur une gamme de températures allant de 15 °C à 95 °C avec un pas de 1 °C. Les données sont traitées avec le logiciel nanotemper. La DLS est mesurée à 20 °C.

#### *3.3.1.4 Formation du complexe alpharep-NS4*

Les alpharep sont concentrées à 500 µM. La fraction de NS4 concentrée à 223 µM provient d'une dialyse dans un tampon sans DDM. Le complexe est formé en mélangeant les alphareps avec NS4 dans un ratio molaire de 1,3 moles d'alpharep pour 1 mole de NS4 (30 µl d'alpharep dans 50 µl de NS4). Les deux partenaires ne sont pas préalablement équilibrés dans le même tampon et le DDM est donc dilué. Le mélange est incubé 10min dans la glace puis 60 µl du mélange est injecté en FPLC Superose 6 increase équilibrée dans le tampon NS4 composé de 0,04 % de DDM, 10mM MES pH6,5 150mM NaCl, et 0,5mM TCEP.

### Profil élution FPLC sur Superose 6 increase



**Figure 3.3.1.4-1.** Profil d'élution du mélange NS4 alpharep en FPLC sur une colonne superose 6 increase.

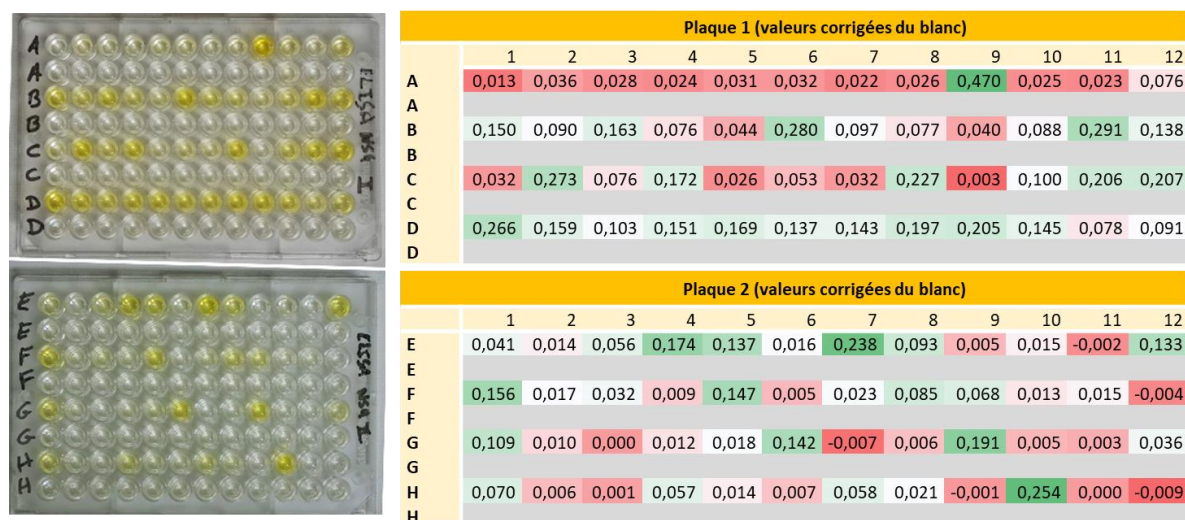
#### 3.3.1.5 Préparation des grilles de cryo-EM et collecte des données

Les grilles sont traitées par plasma pendant 15sec à 22mA. Les grilles avec film de graphène additionnel (UTC-Quantifoil) sont traitées 10sec à 10mA. La vitrification est faite à l'aide d'un vitrobot avec une blotting force à -1 et un blotting time de 4 ou 7 secondes et 1.5  $\mu$ l d'échantillon déposé sur la grille à 20 °C et 100 % d'humidité. Les données sont collectées sur un Glacios équipé d'un filtre à énergie et d'une Falcon 4. La dose totale est de 50e/ $\text{\AA}^2$  sur 567 sub-frames et la taille de pixel de 0,71 $\text{\AA}$ /pixel.

## 3.3.2 Résultats

### 3.3.2.1 Résultats de la sélection des alphareps

Après deux tentatives de révélation par ELISA, nous avons réussi une excellente sélection. Sur la plaque ELISA, une alternance de coloration est visible d'une alpharep à une autre. L'intensité de la coloration est liée à l'affinité de l'alpharep pour NS4, ce qui implique un nombre plus important de phages et donc une meilleure réponse (figure 3.3.2.1-1). Nous avons 7 clones dont l'intensité est supérieure à 0,23 (A9, B6, B11, C2, D1, E7, H10) dont un particulièrement élevé à 0,470 (A9). Nous sélectionnons tous les clones ayant une intensité corrigée du blanc supérieure à 0,1 (50 clones) pour être séquencés (tableau 3.3.2.1-1).



**Figure 3.3.2.1-1.** Résultats de la sélection alpharep. À gauche l'image brute des plaques obtenue par révélation ELISA. À droite les mesures de l'absorbance à 450nm des plaques corrigées du blanc. La correction du blanc est égale à la soustraction de la valeur du puits non spécifique (ligne du bas) à celle du puits spécifique (ligne du haut).

Le séquençage des 50 clones révèle 8 faux positifs correspondant à des mauvaises séquences ou un décalage des phases de lecture (B1, B11, C1, C11, D3, D4, E3, F5). Les neuf séquences correspondantes à des alpharep ayant plus de 8 unités répétées sont incomplètes mais laissent présager au moins 4 séquences uniques. Nous avons donc sélectionné sept séquences selon leurs nombres d'unités répétées (n) et leur intensité lors du test ELISA à savoir, B7(n=2), B3(n=3), C4(n=4), E8(n=6), C12(n=7), A9(n=8), D1(n>9) que nous avons cloné dans le plasmide d'expression. Parmi ces sept nous avons décidé de préparer trois alpharep pour l'expression. La première est C4 qui a été sélectionnée 17 fois et a la taille optimale pour la stabilité et le repliement d'une alpharep avec quatre motifs répétés soit 25,1 kDa. La seconde est A9 car c'est la plus intense du test ELISA et qu'elle est l'une des plus grandes alpharep avec 8 unités répétées soit une taille estimée à 38,8kDa à partir de la séquence. Enfin nous avons décidé de préparer B3 pour sa redondance (2 fois) mais surtout sa petite taille de trois motifs répétés dont nous espérons une fixation différente pour une utilisation en cristallographie.

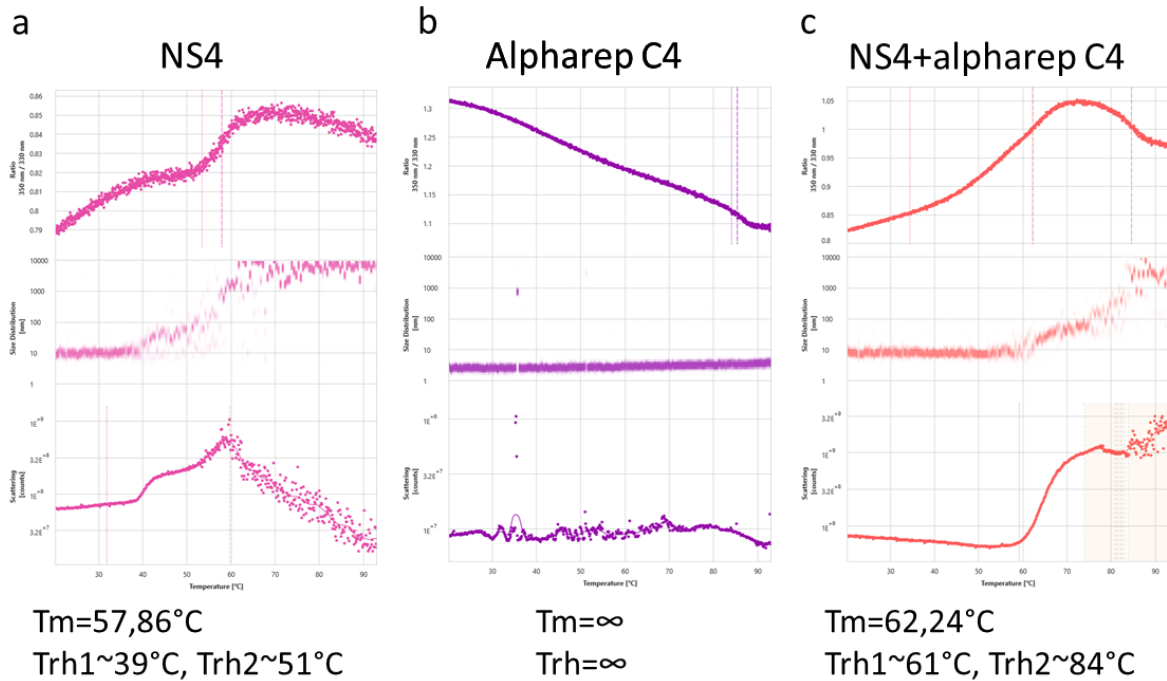
Plan de plaque envoyée au séquençage

	1	2	3	4	5	6	7	8	9	10	11	12	Unités répétées	Nombre de clones	Séquences uniques
A	A9	A12	<b>B1</b>	B2	B3	B4	B6	B7	B8	B10	<b>B11</b>	B12	N < 3	2	2
B	C2	C3	C4	C8	<b>C10</b>	<b>C11</b>	C12	D1	D2	<b>D3</b>	<b>D4</b>	D5	N = 3	4	3
C	D6	D7	D8	D9	D10	D12	E1	<b>E3</b>	E4	E5	E7	E8	N = 4	18	2
D	E12	F1	F3	<b>F5</b>	F8	F9	G1	G6	G9	G12	H1	H4	N = 6	2	2
E	H7	H10											N = 7	1	1
F													N = 8	6	4
G													N ≥ 8	9	~4
H													Total	42	<b>18</b>

**Tableau 3.3.2.1-1** À gauche, le plan de la plaque envoyée au séquençage. Les couleurs indiquent quelques intensités remarquables sur la plaque matrice. En jaune sont les cases correspondantes aux clones ayant une intensité <0,1, en bleu sont ceux ayant une intensité proche de 0,1, en blanc sont ceux dont l'intensité est entre 0,1 et 0,23. Les clones en vert sont les clones dont l'intensité dépasse 0,23. En gras les clones dont le séquençage a révélé une mauvaise séquence. À droite le tableau qui récapitule les résultats du séquençage.

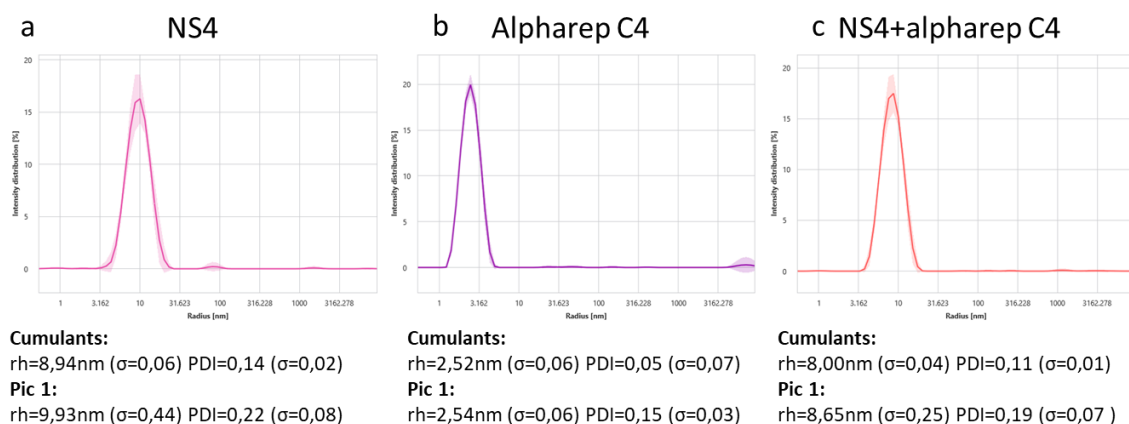
### 3.3.2.2 Analyse biophysique des alphareps avec et sans NS4

Seule l'alpharep C4 a été exprimée et purifiée jusqu'ici. Chacun des partenaires est caractérisé par nano-DSF et DLS. L'alpharep seule est particulièrement stable en nano-DSF. Elle ne donne aucun signe de dénaturation ni d'agrégation sur l'ensemble du gradient de température. Pour ce lot de protéine NS4, la température de fusion est de 57,86 °C et le profil d'agrégation est en deux parties, avec une première augmentation du signal de diffusion et de distribution de taille à 39 °C et un second à 51 °C qui semble indiquer un fort signal d'agrégation. Une fois les deux protéines mélangées ensemble, la température de fusion de NS4 monte à 62 °C. Les températures des variations de diffusion et de distribution de taille augmentent également pour passer à 61 °C et 84 °C. Ces indications sont le signe d'une interaction entre NS4 et l'alpharep qui induit une stabilisation de NS4 (Figure 3.3.2.2-1). Il est intéressant de noter la disparition du signal associé aux micelles de DDM identifié dans le chapitre précédent (figure 3.1.2-1b). En effet ce lot de NS4 a été préalablement dialysé dans un tampon sans DDM ce qui a pour effet de réduire la concentration de DDM. Cet élément suggère que la dialyse a réduit la concentration de détergent, ce qui est peut-être à l'origine de cette différence entre les profils de distribution de taille des figures 3.1.2-1b et 3.3.2.2-1a.



**Figure 3.3.2.2-1.** Spectre de nano-DSF obtenu pour NS4 seule, l'alpharep seule et le mélange NS4 alpharep. La première courbe correspond au ratio 350/330 nm, la seconde montre les mesures DLS, la dernière le signal de diffusion.  $T_m$ = température de fusion,  $Trh$ = température de saut de diffusion.

Les mêmes échantillons sont également analysés par DLS. NS4 seul à un  $rh_{cumulants} \sim 8,9$  nm ( $\pm 0,06$ ) et un indice de polydispersité moyen de 0,14 ( $\pm 0,02$ ). L'alpharep a un  $rh_{cumulants} \sim 2,52$  nm et un indice de polydispersité de 0,05 ( $\pm 0,07$ ). Une fois assemblé/mélangé et avant le passage en FPLC, le  $rh_{cumulants}$  chute légèrement à 8 nm ( $\pm 0,04$ ), l'indice de polydispersité varie faiblement et vaut 0,11 ( $\pm 0,01$ ). Le mélange des alphareps avec NS4 avant le passage en FPLC n'entraîne qu'une faible variation du rayon hydrodynamique et de l'indice de polydispersité. La mesure DLS ne peut pas être une indication fiable de l'assemblage et la diminution peut être induite par l'excédent d'alpharep qui aura tendance à diminuer le coefficient de diffusion et donc le rayon hydrodynamique.

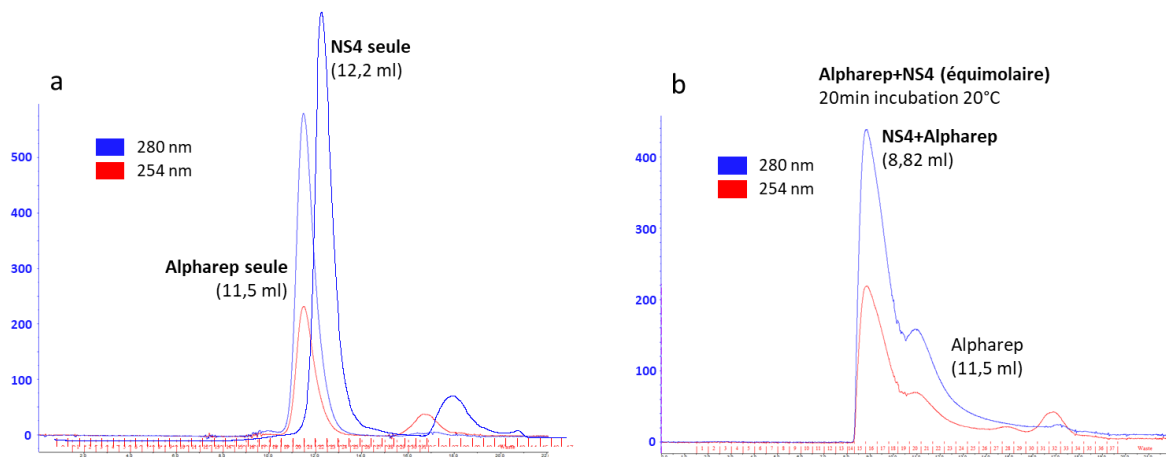


**Figure 3.3.2.2-2.** Mesures DLS de chaque échantillon séparé et du mélange des deux protéines.  $T_m$ = température de fusion, rh= rayon hydrodynamique, PDI = indice de polydispersité.

### 3.3.2.3 Elution du complexe NS4-Alpharep

La capacité de fixation de l'alpharep C4 avec NS4 a été vérifiée sur une S200 analytique. Pour cela trois injections distinctes ont été réalisées, une avec l'alpharep seule, une avec NS4 seule et une dernière avec un mélange équimolaire des deux. Pour chaque injection des profils différents ont été obtenus montrant une fixation de l'alpharep sur NS4. Le contrôle de la formation du complexe est facilité par deux points. Le premier étant le faible signal UV de la protéine NS4 par rapport à l'alpharep. NS4 contient un tryptophane avec un coefficient d'extinction molaire prédit à  $14440 \text{ M}^{-1}\text{cm}^{-1}$  (0,71 Abs pour 1 g/l) et C4 porte 3 tryptophanes avec un coefficient d'extinction molaire prédit à  $22460 \text{ M}^{-1}\text{cm}^{-1}$  (0,9 Abs pour 1 g/l pour C4). Le second est la résolution du pic d'éluion de l'alpharep, dont le profil fin est facile à identifier. Le profil d'éluion du complexe NS4/Alpharep en akta sur une S200 se caractérise par le déplacement du signal de l'alpharep vers un temps de rétention plus court, qui correspondrait à la formation du complexe NS4-Alpharep (Figure 3.3.2.3-1).





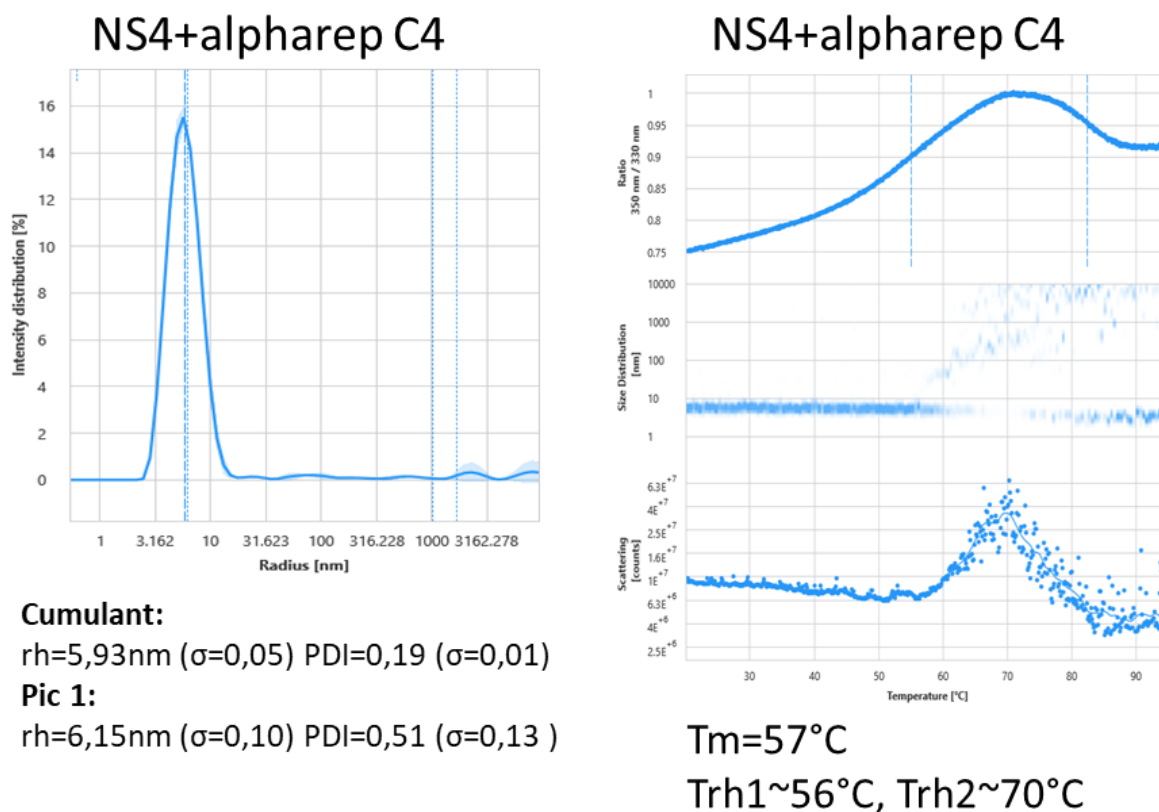
**Figure 3.3.2.3-1.** Expérience réalisée par Adrien Royet. Comparaison des profils d'élutions en chromatographie d'exclusion sur une colonne S200 de 24ml. À gauche (a) les profils d'élution de l'alpharep seule et de NS4 seule dans le tampon standard de NS4 (10 mM MES pH6.5, 150 mM NaCl, 0,5 mM TCEP et 0,04% DDM) à droite le profil d'élution du mélange NS4+Alpharep sur la même colonne dans les mêmes conditions. Le premier pic correspond à un complexe NS4+alpharep, le second pic correspond à l'excès d'alpharep. Les courbes en bleu correspondent à l'absorbance à 280 nm et en rouge à 254 nm.

Les fractions 3 et 4 sont analysées par UV, nano-DSF et nano DLS (figure 3.3.2.3-2). En premier lieu nous retrouvons le signal typique de la présence de DDM à forte concentration dans les mesures DLS au cours du gradient de température. Celui-ci se caractérise par deux populations distinctes, dont une avec un rayon de 5 nm, qui pourrait être les micelles de DDM comme sur la figure 3.1.2-1b, et l'autre à très haute taille. De plus en DLS classique, nous observons une diminution du rayon hydrodynamique global à 5,93 nm ( $\pm 0,05$ ). C'est plus bas que celui observé avant l'élution dans un tampon contenant moins de détergent (figure 3.3.2.2-2c). Du côté de la nano-DSF, la température de fusion est d'environ 58 à 60 °C très similaire à celle observée sans alpharep dans les mêmes conditions. En revanche les signes d'agrégation se manifestent nettement plus tard à 56 °C pour le premier évènement et 70 °C pour le second. Ce résultat est paradoxal du point de vue de la température de fusion qui ne semble pas s'améliorer après l'ajout de l'alpharep. En revanche il y a une légère amélioration de l'agrégation qui semble arriver plus tardivement. Les données indiquent donc un rôle stabilisant du détergent prépondérant à celui de l'alpharep.

Si nous comparons les trois profils nano-DSF complet de NS4, en condition réduite en détergent (figure 3.3.2.2-1a), celui équivalent après l'ajout de l'alpharep (figure 3.3.2.2-1c) et celui de NS4 seule dans 0,04 % de DDM (figure 1.2-1b), nous remarquons que l'alpharep permet une amélioration du  $T_m$  ce qui présage une meilleure stabilité dans cette condition. Enfin, la comparaison des données DLS d'un échantillon fraîchement sorti de FPLC obtenues avec l'alpharep dans 0,04 % de DDM (figure 3.3.2.3-2) ou sans alpharep (figure 1.2-1), montre une légère augmentation du rayon hydrodynamique en présence de l'alpharep qui passe de 5,21 nm ( $\pm 0,02$ ) à 5,93 nm ( $\pm 0,05$ ). À titre de comparaison, le complexe HTP du chapitre

précédent avait un rayon hydrodynamique par DLS de 6,15nm (Figure 1.2-1) donc relativement proche.

## Mesures après FPLC

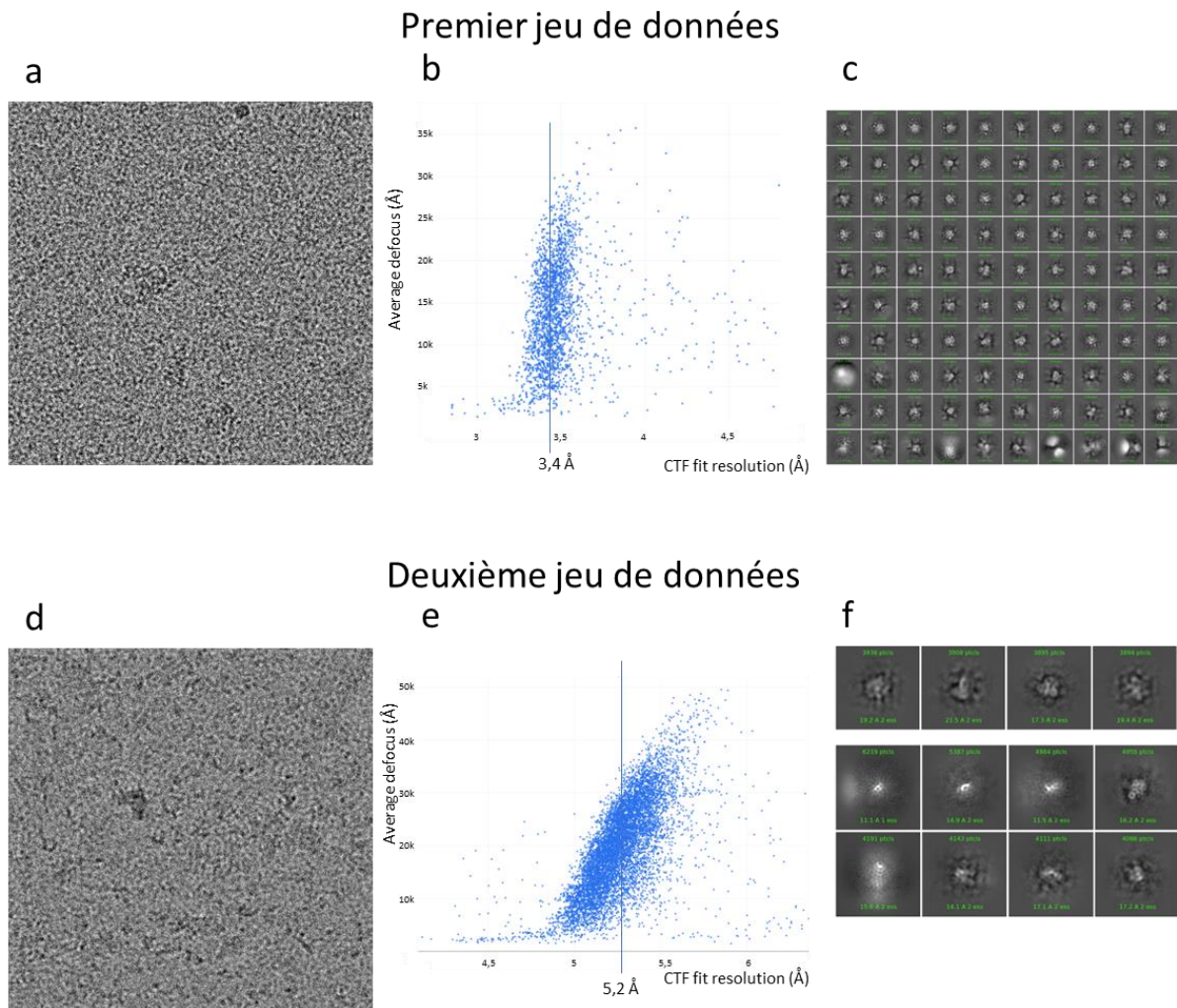


**Figure 3.3.2.3-2.** Mesure de DLS et nano-DSF du complexe NS4-Alpharep après l'élution en FPLC sur une colonne superose 6 increase dans un tampon contenant 0,04 % de DDM.

### 3.3.2.4 Premier essai cryo-EM SPA

Un premier essai de microscopie électronique a été fait avec les alpharep. Deux jeux ont été collectés sur deux grilles UTC-Quantifoil préparées avec des paramètres de Vitrobot différents. Le traitement révèle une bonne qualité de la grille par la résolution moyenne via CTF, qui se situe autour des 3,4 Å pour le premier jeu et 5,2 Å pour le deuxième, et sur les images qui ont une bonne glace et peu de contamination. Il semble possible de discerner le signal des particules (figure 3.3.2.3-3). La résolution estimée par CTF sur ce type de grilles est biaisée par la présence du graphène continu. L'aspect de la distribution du plot CTF vs Defocus de la première grille suggère une glace très fine ou l'absence de glace, tandis que sur le deuxième jeu de données, la distribution est plus caractéristique de la diminution de la résolution à plus haute défocalisation. Malgré ce constat de bonne qualité des jeux, il est impossible d'identifier un motif clair par classification 2D (figure 3.3.2.3-3). La première hypothèse est que le complexe a été préparé avec trop de détergent, et l'objet étant petit, son signal est noyé dans le bruit des micelles. La deuxième hypothèse est que le complexe réagit mal au Vitrobot et est

absorbé lors de l'étape de "blotting". Enfin, il est aussi possible que l'objet formé n'adopte pas une conformation uniforme et soit donc très hétérogène, ce qui réduirait drastiquement les chances d'être observé en classification 2D.



**Figure 3.3.2.3-3.** À gauche (**a, d**) i, exemple d'une image pour chaque jeu NS4-alphaRep. Au centre (**b,e**) les graphiques qui montrent le plot de la résolution estimée par CTF (CTFfit) en fonction de la défocalisation. À droite (**c,f**) exemple de classes 2D.

### 3.3.3 Conclusion et perspectives : un catalogue d'alpharep diversifié disponible

Les alphareps ont pu être sélectionnées avec un haut taux de réussite. Nous disposons à présent d'un catalogue diversifié de 18 alphareps uniques de 2 à 8 unités répétées soit de 13kDa à 32kDa. Le choix s'est porté sur C4 pour les premiers essais d'interaction car bien qu'elle ne soit pas la plus grande en taille et donc pas la favorite pour la cryo-EM, elle présente un profil idéal pour une alpharep et avait donc de fortes chances de succès. Les grosses alphareps ont la réputation de former des oligomères ce qui pourrait entraver leur capacité de fixation à NS4, mais restent intéressantes pour la cryo-EM.

Nous avons également observé des phénomènes très intéressants liés à l'introduction de l'alpharep. Lors du mélange NS4-Alpharep avant le passage en FPLC, bien que le DDM soit dilué, nous observons une stabilisation du rayon hydrodynamique. Cette étape entraîne la dilution de NS4 à 139  $\mu$ M ainsi que le détergent et le TCEP déjà faiblement concentrés qui sont dilués environ au tiers. Nous avons démontré qu'à cette concentration de NS4, elle est sujette à l'oligomérisation (tableau et figure 1.2-1) or ici ni la DLS ni le profil d'élution FPLC n'indiquent la formation de gros objets de type agrégat. En revanche le maintien d'un rh relativement élevé dans ces mesures s'explique par la condition faiblement réductrice du mélange. Le TCEP présent dans l'échantillon NS4 avait un pouvoir réducteur vieillissant et permettrait la persistance de certains ponts disulfures. Après le passage en FPLC dans une condition fortement réductrice, le rayon hydrodynamique diminue (figure 3.2.3-2). En effet les conditions réductrices, la forte concentration en DDM, et la faible concentration de NS4 la pousse vers un état monomérique.

Les données nano-DSF montrent également l'effet stabilisant marqué du DDM et de l'alpharep pour NS4. Le simple ajout de DDM ou d'alpharep dans l'échantillon NS4 pauvre en détergent permet d'augmenter la température de fusion de NS4 de +10 °C. Couplée à l'observation DLS, il est possible que l'alpharep C4 couvre une partie de la région hydrophobe de NS4 et favorise sa solubilisation. Ainsi à l'avenir, NS4 et l'alpharep devront être co-éluées en FPLC dans un tampon contenant une concentration de DDM de 0,017 %. Ceci devrait permettre d'obtenir de plus belles images en cryo-EM en réduisant le signal associé aux micelles de DDM.

Dans les conditions actuelles, nous pouvons espérer un complexe d'une taille minimale approximative de 46kDa composé de NS4<sub>FL</sub> monomérique (21kDa), l'alpharep C4 (25kDa) en plus du DDM (~40kDa). Il est fort probable que NS4 soit sous une forme multimérique même en condition réductrice, et donc que l'objet soit en réalité plus gros que cette estimation. L'objet est donc en l'état encore relativement petit pour se placer dans des conditions idéales pour atteindre la haute résolution. La première tentative de cryo-EM a d'ailleurs montré que les logiciels ne parvenaient pas à extraire un signal NS4-Alpharep probablement à cause du bruit provoqué par les micelles de DDM.

La première option la plus accessible pour apporter une première amélioration sera de tester

les autres alphareps plus grandes, notamment A9 (32kDa), qui pourra améliorer le signal des images de microscopie électronique. La seconde option serait de tester la fixation simultanée de différents alphareps. La troisième piste serait d'utiliser le mutant NS4<sub>C160A</sub>, dont la cystéine 160 de la zone non structurée a été remplacé par une alanine. Par cette mutation nous aimerions à la fois permettre la formation de dimères de NS4 stables par cross-linking à partir des cystéines 103 et/ou 112, tout en empêchant l'oligomérisation de haut poids moléculaire. Ce genre d'assemblage pourrait aboutir à un complexe de 96kDa sans comptabiliser le signal du DDM. Nous serions alors dans une gamme de taille bien plus favorable à la Cryo-EM. Dans le même esprit, l'obtention d'alphareps dirigées contre NS4 permet d'envisager des outils originaux développés au sein de l'I2BC par l'équipe Modélisation et Ingénierie des Protéines. L'un des projets consiste à fusionner les alphareps à une grosse protéine telle que la beta-galactosidase sur lequel viendrait se fixer la protéine NS4 et qui formerait un objet idéal pour la microscopie électronique.

Une autre application pragmatique serait d'étudier l'effets des alphareps lors des manipulations *in-vitro* menées au sein de l'équipe sur les déformations de liposomes. Une fois fusionnées avec un marqueur fluorescent, elles pourraient aussi avoir la même application que les mAbs présentées dans la figure 2.1.2.2-2 avec l'avantage de coûter moins cher et d'être plus accessibles. Mais l'une des grandes forces des alphareps réside dans leur capacité à pouvoir être exprimées fonctionnelles dans les cellules vivantes. Ceci ouvre une myriade de possibilités pour les expériences *in-vivo* (Chevrel et al. 2015; Valerio-Lepiniec et al. 2015). Avec un tel système nous pouvons envisager un marquage dans les cellules vivantes pour simplement tracer la protéine ou mesurer les éventuelles activités inhibitrices. À la suite de quoi il serait possible de combiner l'information structurale avec une activité *in-vivo*. Leur expression pouvant être médiée par une induction contrôlée, il serait aussi possible d'étudier leur effet à différents moments du processus infectieux.

Si nous nous projetons encore plus loin, nous pouvons tout à fait imaginer une application thérapeutique des alphareps. En effet, les alphareps et les DARPins ou de manière générale les protéines à répétition ankyrines sont déjà fortement étudiées pour des applications thérapeutiques (Nangola et al. 2012; Sakkhachornphop et al. 2018; Gebauer and Skerra 2020; Walser et al. 2022). Avec NS4 nous sommes dans un cas typique où il n'existe pas de traitement pour lutter contre les infections au norovirus. NS4 étant essentielle à la multiplication virale cela en fait une cible thérapeutique de premier choix. L'application de ces protéines recombinantes est encore une niche mais certaines ont déjà montré leur efficacité en essai clinique (Boersma 2018; Stumpp et al. 2020; ACTIV-3/TICO et al. 2022) et nous pouvons envisager qu'elles s'implanteront dans le paysage pharmaceutique avec peut-être une application similaire à celle des Fabs, voire même, si elles évoluent, se rapprocher des mAbs. Les principaux avantages des alphareps sont leur faible coût de production et leur petite taille permettant d'avoir une bien meilleure biodisponibilité que les mAbs (Boersma 2018). Cependant l'utilisation thérapeutique d'alpharep se heurte à plusieurs limitations d'ordre pharmacocinétiques. Leur petite taille est associée à un temps de demi-vie

particulièrement court ainsi qu'une clairance élevée. Ces propriétés sont utiles pour les applications en imagerie médicale mais limitantes pour les applications thérapeutiques. De même, les anticorps établissent un lien direct entre l'immunité humorale et le système immunitaire inné par le Fc et le recrutement du complément contribuant largement à leur efficacité. Cependant, dans le cas d'une application pour une neutralisation de NS4 il faudrait trouver un moyen d'internaliser les alphareps dans les cellules infectées, ce qui représente un gros projet d'optimisation.

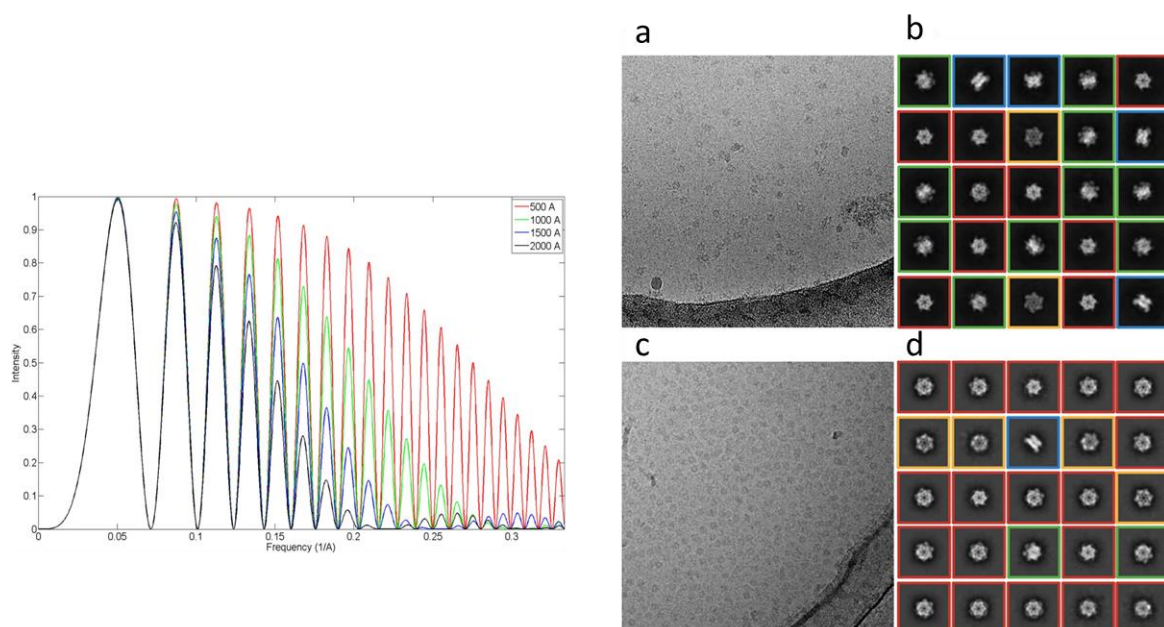
Il reste encore beaucoup de mystères autour de la protéine NS4, mais les résultats présentés ici ont apporté des éléments qui ont ouvert de nouvelles perspectives. Bien que de grandes étapes aient été franchies, il faut encore quelques efforts avant de parvenir à la résolution de la structure de NS4 par cryo-EM. Mais NS4 se purifie bien et en quantité raisonnable, ce qui permet de multiplier les essais dans des délais relativement courts. Je pense que quelques mois supplémentaires m'auraient permis l'obtention de premières classes 2D si ce n'est un volume convaincant.



## 4 Discussion Générale

Il existe de nombreux outils qui peuvent servir à mieux diagnostiquer la qualité d'un jeu de données. Nous avons vu qu'au cours du processus plusieurs paramètres peuvent être à l'origine de l'obtention d'une carte à basse résolution. Les trois principales limitations qui ont été rencontrées sont la qualité de la grille, l'épaisseur de la glace et le comportement de l'échantillon sur la grille.

L'épaisseur de la glace est une grande limitation en cryo-EM. Le problème, c'est qu'une variation de quelque dizaine de nanomètres impacte drastiquement la résolution maximale de l'image produite (Wu et al. 2015; Neselu et al. 2023). Cet effet est directement visible sur la CTF et les anneaux de Thong (figure 4-1). L'épaisseur de glace est un élément qui peut également impacter la distribution et l'orientation des particules sur la grille (Vinothkumar and Henderson 2016; Kampjut et al. 2021; Huntington et al. 2022). Nous nous sommes d'ailleurs interrogés sur ce point pour le complexe HTP, la glace épaisse est-elle favorable à la bonne distribution ? Ou l'inverse ? Une glace trop fine est-elle à l'origine du phénomène de "clustering"/agglomération de HTP ? Difficile à dire sans des mesures effectives et systématique associées aux images. C'est pour cela que de récentes publications proposent des stratégies pour mesurer systématiquement l'épaisseur de la glace tout au long de la collecte (Rice et al. 2018; Brown and Hanssen 2022). Je pense qu'il serait bénéfique de déployer ces logiciels sur les plateformes. Cela apportera une information que je pense être parfois cruciale pour conclure sur le véritable facteur limitant d'un jeu de données.



**Figure 4-1.** À gauche, graphique repris de la revue *Microscopy (Oxf)*, Volume 65, Issue 1 (Wu et al. 2015). Les images **a**, **b**, **c**, **d** proviennent de Huntington et al. 2022 (Huntington et al. 2022). Le graphique des différentes CTF montre comment l'épaisseur de la glace impacte la résolution théorique maximale de l'image. Images enregistrées à  $-1 \mu\text{m}$  de défocalisation. Les



*CTF sont déterminées en fonction de différentes épaisseurs de la glace : 50 nm en rouge, 100 nm en vert, 150 en bleu, 200 nm en noir. Les images à droite montrent que des orientations plus diversifiées sont observées dans une glace épaisse (a,b) que dans une glace plus fine (c,d). Les classes 2D sont encadrées de différentes couleurs selon l'orientation correspondante.*

## **L'impact environnemental de la Cryo-EM**

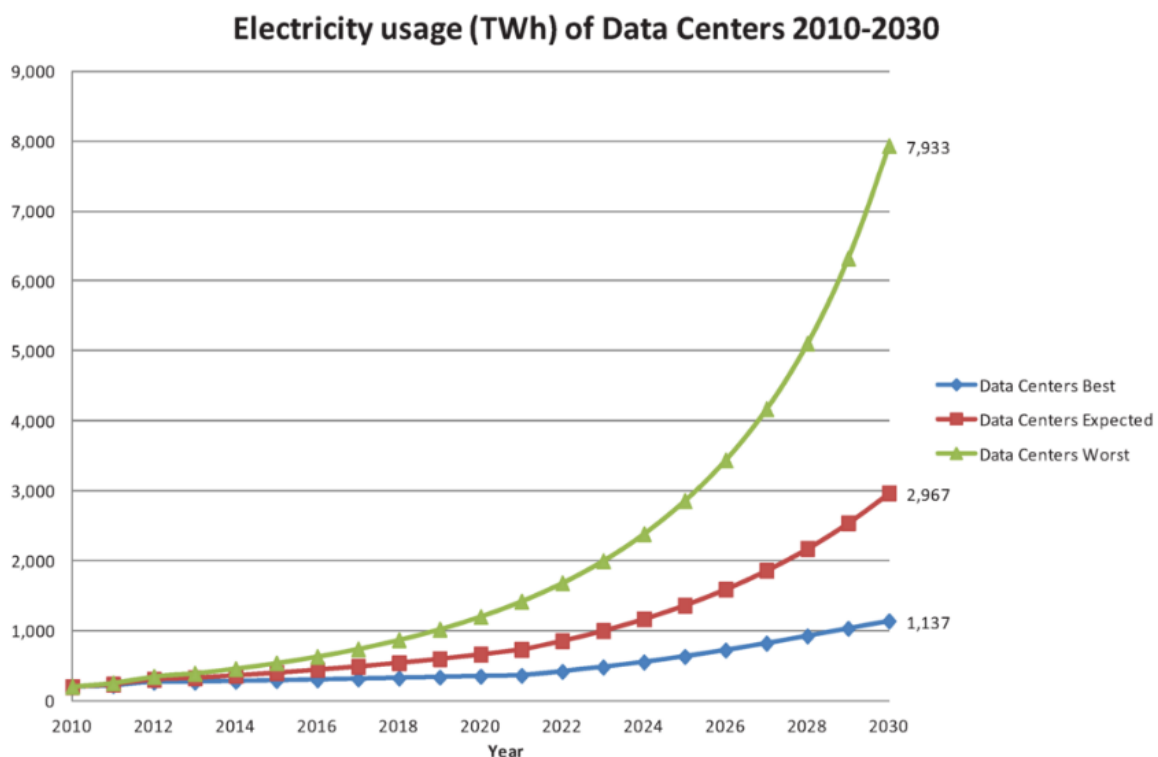
J'aimerais aussi appuyer sur la pertinence des critères qualité préliminaire à la collecte de données. L'accès à un microscope n'est aujourd'hui plus un facteur limitant. Les nouvelles caméras permettent de prendre jusqu'à 500 à 600 films à l'heure, et avec les systèmes "multigrids" les nouveaux microscopes sont de redoutables machines à générer des données. De même, le traitement de ces données est très accessible grâce à des logiciels optimisés et très rapides à prendre en main. Une station équipée d'une ou deux cartes graphiques récentes et d'un processeur de 32 cœurs d'une valeur de 10000€ suffit pour assurer le traitement de 2-3 jeux de données en parallèles. Cette relative accessibilité a contribué à l'engouement envers la cryo-EM. Cependant, il faut être méfiant quant à une utilisation prématurée dans un projet avec un espoir de "diagnostic". Une grosse limitation vient de la taille des fichiers générés pendant une collecte. Étendons-nous sur le sujet, souvent un jeu de données fait entre 5000 et 10000 films 4k de 3-5 secondes chacun enregistré à 320 images par seconde. Soit mis bout à bout l'équivalent d'un film de 5 à 11 heures, Spielberg n'a qu'à bien se tenir. Présenté ainsi on comprend mieux l'enclume qu'on aura à notre charge. En 2022 le nouveau format eer a permis de réduire la taille des films tout en conservant (voir améliorant) la résolution. Par exemple une collecte de 5000 films pèse environ ~2,3To de données contre ~2,6To en format tiff. Bien qu'encourageante, cette compression n'est pas révolutionnaire et elle ne permet qu'une timide amélioration. Enfin, au poids de la collecte, s'ajoute celui du traitement des données qui peut facilement faire doubler le volume de données.

On comprend donc le problème émergent du stockage des données qui touche les instituts et industriels qui s'équipent de microscopes électroniques. En pratique a opté pour un archivage rapide sur bande, qui permet de désengorger les clusters déjà saturés. À l'institut Pasteur depuis 2023, les jeux collectés par les utilisateurs extérieurs ne sont plus conservés. Ailleurs comme au biozentrum de l'université de Basel, les jeux ne sont conservés qu'une semaine et l'archivage est laissé à la responsabilité de l'utilisateur. J'ai observé aussi fréquemment un archivage en local dans des disques durs achetés par les équipes de recherche. Enfin Sanofi a opté pour une location de serveur distant Amazon. Bien que le prix du stockage soit en baisse constante, il reste encore cher et il ajoute une ligne, à la déjà lourde facture de la cryo-EM.

À l'heure où les bouleversements climatiques chamboulent dangereusement et de manière inquiétante le monde, le stockage de l'information prend une part importante dans l'émission des gaz à effet de serre. L'arrivée d'un microscope s'accompagne généralement du renforcement des services informatiques existants ou de l'émergence de mini "data-center"

pour gérer les données. Or un regard sur le monde montre que les data center consomment près de 3 % de l'électricité mondiale et certaines prévisions tendent vers un 10 % à l'horizon 2030 (figure 4-2) (Andrae and Edler 2015). Si à cela, on considère l'extraction des matériaux, la fabrication la distribution et la fin de vie des équipements qui compose les matériaux de stockage de l'information cela donne vite le vertige quant à leur impact négatif sur l'environnement. Même si la cryo-EM n'est pas l'acteur majoritaire de cette problématique, elle y contribue. Pour remédier à ce problème la solution de l'archivage des films et des résultats de traitement est efficace mais, les données sont "dormantes" et donc inaccessibles. Dans les faits, des jeux peuvent rester actifs une ou deux années, parfois plus comme c'est le cas dans ma thèse. Cela se justifie par le fait que des nouveaux logiciels et programmes voient le jour continuellement et permettent de nouvelles analyses qui peuvent améliorer l'analyse et apporter de nouvelles informations sur l'échantillon. C'est le cas de ceux que nous avons présenté dans le manuscrit avec 3DFlex et MDSPACE par exemple, ou l'affinement localisé avant eux.

Je pense qu'il est raisonnable et pertinent de s'interroger sur une approche plus écoresponsable de la recherche, c'est aussi cette idée qui est défendue par la réflexion méthodologique développée dans ce manuscrit. Raisonner sur une approche optimisée pour la cryo-EM limite son coût et son impact environnemental.



**Figure 4-2.** Graphique issu de Andrae Anders et Elter Tomas 2015 (Andrae and Edler 2015). Courbes montrant l'évolution de la consommation électrique des data center dans le monde et la tendance pour les années à venir. En vert le scénario d'une explosion plus importante que prévu, en bleu une estimation minorée et en orange l'estimation attendue par l'étude.



## 5 Références

- ACTIV-3/TICO, Barkauskas C, [...], Williams R (2022) Efficacy and Safety of Ensovibep for Adults Hospitalized With COVID-19. *Ann Intern Med* 175(9):M22-1503. <https://doi.org/10.7326/m22-1503>
- Afonine PV, Poon BK, Read RJ, Sobolev OV, Terwilliger TC, Urzhumtsev A, Adams PD (2018) Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr Sect D* 74(6):531–544. <https://doi.org/10.1107/s2059798318006551>
- Andrade MA, Bork P (1995) HEAT repeats in the Huntington’s disease protein. *Nat Genet* 11(2):115–116. <https://doi.org/10.1038/ng1095-115>
- Andrade MA, Perez-Iratxeta C, Ponting CP (2001a) Protein Repeats: Structures, Functions, and Evolution. *J Struct Biol* 134(2–3):117–131. <https://doi.org/10.1006/jsbi.2001.4392>
- Andrade MA, Petosa C, O’Donoghue SI, Müller CW, Bork P (2001b) Comparison of ARM and HEAT protein repeats 11 Edited by P. E. Wright. *J Mol Biol* 309(1):1–18. <https://doi.org/10.1006/jmbi.2001.4624>
- Andrae ASG, Edler T (2015) On Global Electricity Usage of Communication Technology: Trends to 2030. *Challenges* 6(1):117–157. <https://doi.org/10.3390/challe6010117>
- AntibodySociety (2023) Antibody therapeutics approved or in regulatory review in the EU or US - The Antibody Society. <https://www.antibodysociety.org/resources/approved-antibodies/>. Accessed 18 Sep 2023
- Asarnow D, Palovcak E, Cheng & Y (2019) UCSF pyem. Zenodo. <https://doi.org/10.5281/zenodo.3576630>
- Austin CD, Mazière AMD, Pisacane PI, Dijk SM van, Eigenbrot C, Sliwkowski MX, Klumperman J, Scheller RH (2004) Endocytosis and Sorting of ErbB2 and the Site of Action of Cancer Therapeutics Trastuzumab and Geldanamycin. *Mol Biol Cell* 15(12):5268–5282. <https://doi.org/10.1091/mbc.e04-07-0591>
- Bai X, Sun P, Wang X, Long C, Liao S, Dang S, Zhuang S, Du Y, Zhang X, Li N, He K, Zhang Z (2023) Structure and dynamics of the EGFR/HER2 heterodimer. *Cell Discov* 9(1):18. <https://doi.org/10.1038/s41421-023-00523-5>
- Baker LA, Rubinstein JL (2010) Chapter Fifteen Radiation Damage in Electron Cryomicroscopy. *Methods Enzym* 481:371–388. [https://doi.org/10.1016/s0076-6879\(10\)81015-8](https://doi.org/10.1016/s0076-6879(10)81015-8)
- Baker MR, Fan G, Serysheva II (2014) Single-particle cryo-EM of the ryanodine receptor channel. *Eur J Transl Myol* 25(1):35–48. <https://doi.org/10.4081/bam.2015.1.35>

- Ben-Kasus T, Schechter B, Lavi S, Yarden Y, Sela M (2009) Persistent elimination of ErbB-2/HER2-overexpressing tumors using combinations of monoclonal antibodies: Relevance of receptor endocytosis. *Proc Natl Acad Sci* 106(9):3294–3299. <https://doi.org/10.1073/pnas.0812059106>
- Bepler T, Morin A, Rapp M, Brasch J, Shapiro L, Noble AJ, Berger B (2019) Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods* 16(11):1153–1160. <https://doi.org/10.1038/s41592-019-0575-8>
- Binz HK, Stumpp MT, Forrer P, Amstutz P, Plückthun A (2003) Designing Repeat Proteins: Well-expressed, Soluble and Stable Proteins from Combinatorial Libraries of Consensus Ankyrin Repeat Proteins. *J Mol Biol* 332(2):489–503. [https://doi.org/10.1016/s0022-2836\(03\)00896-9](https://doi.org/10.1016/s0022-2836(03)00896-9)
- Blumenthal GM, Scher NS, Cortazar P, Chattopadhyay S, Tang S, Song P, Liu Q, Ringgold K, Pilaro AM, Tilley A, King KE, Graham L, Rellahan BL, Weinberg WC, Chi B, Thomas C, Hughes P, Ibrahim A, Justice R, Pazdur R (2013) First FDA Approval of Dual Anti-HER2 Regimen: Pertuzumab in Combination with Trastuzumab and Docetaxel for HER2-Positive Metastatic Breast Cancer. *Clin Cancer Res* 19(18):4911–4916. <https://doi.org/10.1158/1078-0432.ccr-13-1212>
- Boersma YL (2018) Protein Scaffolds, Design, Synthesis, and Applications. *Methods Mol Biol* 1798:307–327. [https://doi.org/10.1007/978-1-4939-7893-9\\_23](https://doi.org/10.1007/978-1-4939-7893-9_23)
- Bostrom J, Yu S-F, Kan D, Appleton BA, Lee CV, Billeci K, Man W, Peale F, Ross S, Wiesmann C, Fuh G (2009) Variants of the Antibody Herceptin That Interact with HER2 and VEGF at the Antigen Binding Site. *Science* 323(5921):1610–1614. <https://doi.org/10.1126/science.1165480>
- Bouché O, Penault-Llorca F (2010) [HER2 and gastric cancer: a novel therapeutic target for trastuzumab]. *Bull du cancer* 97(12):1429–40. <https://doi.org/10.1684/bdc.2010.1224>
- Brilot AF, Chen JZ, Cheng A, Pan J, Harrison SC, Potter CS, Carragher B, Henderson R, Grigorieff N (2012) Beam-induced motion of vitrified specimen on holey carbon film. *J Struct Biol* 177(3):630–637. <https://doi.org/10.1016/j.jsb.2012.02.003>
- Briney B, Inderbitzin A, Joyce C, Burton DR (2019) Commonality despite exceptional diversity in the baseline human antibody repertoire. *Nature* 566(7744):393–397. <https://doi.org/10.1038/s41586-019-0879-y>
- Brown HG, Hanssen E (2022) MeasureIce: accessible on-the-fly measurement of ice thickness in cryo-electron microscopy. *Commun Biology* 5(1):817. <https://doi.org/10.1038/s42003-022-03698-x>
- Chevrel A, Urvoas A, Sierra-Gallay IL de la, Aumont-Nicaise M, Moutel S, Desmadril M, Perez F, Gautreau A, Tilbeurgh H van, Minard P, Valerio-Lepiniec M (2015) Specific GFP-binding artificial proteins ( $\alpha$ Rep): a new tool for in vitro to live cell applications. *Biosci Rep* 35(4):e00223. <https://doi.org/10.1042/bsr20150080>

- Cho H-S, Mason K, Ramyar KX, Stanley AM, Gabelli SB, Denney DW, Leahy DJ (2003) Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature* 421(6924):756–760. <https://doi.org/10.1038/nature01392>
- D’Imprima E, Kühlbrandt W (2021) Current limitations to high-resolution structure determination by single-particle cryoEM. *Q Rev Biophys* 54:e4. <https://doi.org/10.1017/s0033583521000020>
- Diwanji D, Trenker R, Thaker TM, Wang F, Agard DA, Verba KA, Jura N (2021) Structures of the HER2–HER3–NRG1 $\beta$  complex reveal a dynamic dimer interface. *Nature* 600(7888):339–343. <https://doi.org/10.1038/s41586-021-04084-z>
- Doerflinger SY, Cortese M, Romero-Brey I, Menne Z, Tubiana T, Schenk C, White PA, Bartenschlager R, Bressanelli S, Hansman GS, Lohmann V (2017) Membrane alterations induced by nonstructural proteins of human norovirus. *PLoS Pathog* 13(10):e1006705. <https://doi.org/10.1371/journal.ppat.1006705>
- Einstein A (1905) INVESTIGATIONS ON THE THEORY OF THE BROWNIAN MOVEMENT. DOVER PUBLICATIONS
- El-Gamal MI, Mewafi NH, Abdelmotteleb NE, Emara MA, Tarazi H, Sbenati RM, Madkour MM, Zarai S-O, Shahin AI, Anbar HS (2021) A Review of HER4 (ErbB4) Kinase, Its Impact on Cancer, and Its Inhibitors. *Molecules* 26(23):7376. <https://doi.org/10.3390/molecules26237376>
- EMA EMA (2000) Herceptin | European Medicines Agency. <https://www.ema.europa.eu/en/medicines/human/EPAR/herceptin>. Accessed 25 Apr 2023
- EMA EMA (2013) Perjeta | European Medicines Agency. <https://www.ema.europa.eu/en/medicines/human/EPAR/perjeta>. Accessed 30 May 2023
- Fan R, Shi X, Guo B, Zhao J, Liu J, Quan C, Dong Y, Fan S (2021) The effects of L-arginine on protein stability and DNA binding ability of SaeR, a transcription factor in *Staphylococcus aureus*. *Protein Express Purif* 177:105765. <https://doi.org/10.1016/j.pep.2020.105765>
- Fatemi HE, Hammas N, Idrissi K, Mellas N, Bennani A, Amarti A, Cayre A, Penault-Illorca F, Mesbahi O (2012) [The search for protein overexpression and HER2 gene amplification in gastric cancer by immunohistochemistry and in situ hybridization: experience of the CHU Hassan II of Fez]. *Pan Afr Méd J* 13:79
- FDA (1998) Herceptin (trastuzumab) FDA Approval History - Drugs.com. <https://www.drugs.com/history/herceptin.html>. Accessed 30 May 2023
- FDA F and DA (2012) Perjeta: Uses, Dosage, Side Effects & Warnings - Drugs.com. <https://www.drugs.com/perjeta.html>. Accessed 30 May 2023
- Franken LE, Grünewald K, Boekema EJ, Stuart MCA (2020) A Technical Introduction to Transmission Electron Microscopy for Soft-Matter: Imaging, Possibilities, Choices, and Technical Developments. *Small* 16(14):1906198. <https://doi.org/10.1002/sml.201906198>

- Franklin MC, Carey KD, Vajdos FF, Leahy DJ, Vos AM de, Sliwkowski MX (2004) Insights into ErbB signaling from the structure of the ErbB2-pertuzumab complex. *Cancer Cell* 5(4):317–328. [https://doi.org/10.1016/s1535-6108\(04\)00083-2](https://doi.org/10.1016/s1535-6108(04)00083-2)
- Fu W, Wang Y, Zhang Y, Xiong L, Takeda H, Ding L, Xu Q, He L, Tan W, Bethune AN, Zhou L (2014) Insights into HER2 signaling from step-by-step optimization of anti-HER2 antibodies. *Mabs* 6(4):978–990. <https://doi.org/10.4161/mabs.28786>
- Gebauer M, Skerra A (2020) Engineered Protein Scaffolds as Next-Generation Therapeutics. *Annu Rev Pharmacol Toxicol* 60(1):391–415. <https://doi.org/10.1146/annurev-pharmtox-010818-021118>
- Ghosh R, Narasanna A, Wang SE, Liu S, Chakrabarty A, Balko JM, González-Angulo AM, Mills GB, Penuel E, Winslow J, Sperinde J, Dua R, Pidaparathi S, Mukherjee A, Leitzel K, Kostler WJ, Lipton A, Bates M, Arteaga CL (2011) Trastuzumab Has Preferential Activity against Breast Cancers Driven by HER2 Homodimers. *Cancer Res* 71(5):1871–1882. <https://doi.org/10.1158/0008-5472.can-10-1872>
- Glaeser RM (2016) Chapter Two Specimen Behavior in the Electron Beam. *Methods Enzym* 579:19–50. <https://doi.org/10.1016/bs.mie.2016.04.010>
- Grove TZ, Cortajarena AL, Regan L (2008) Ligand binding by repeat proteins: natural and designed. *Curr Opin Struct Biol* 18(4):507–515. <https://doi.org/10.1016/j.sbi.2008.05.008>
- Guellouz A, Valerio-Lepiniec M, Urvoas A, Chevrel A, Graille M, Fourati-Kammoun Z, Desmadril M, Tilbeurgh H van, Minard P (2013) Selection of Specific Protein Binders for Pre-Defined Targets from an Optimized Library of Artificial Helicoidal Repeat Proteins (alphaRep). *PLoS ONE* 8(8):e71512. <https://doi.org/10.1371/journal.pone.0071512>
- Guo H, Franken E, Deng Y, Benlekbir S, Lezcano GS, Janssen B, Yu L, Ripstein ZA, Tan YZ, Rubinstein JL (2020) Electron-event representation data enable efficient cryoEM file storage with full preservation of spatial and temporal resolution. *IUCrJ* 7(Pt 5):860–869. <https://doi.org/10.1107/s205225252000929x>
- Habersetzer J, Debbah M, Fogeron M-L, Böckmann A, Bressanelli S, Fieulaine S (2020) In vitro translation of virally-encoded replication polyproteins to recapitulate polyprotein maturation processes. *Protein Expr Purif* 175:105694. <https://doi.org/10.1016/j.pep.2020.105694>
- Hamitouche I, Jonic S (2022) DeepHEMNMA: ResNet-based hybrid analysis of continuous conformational heterogeneity in cryo-EM single particle images. *Front Mol Biosci* 9:965645. <https://doi.org/10.3389/fmolb.2022.965645>
- Hanske J, Sadian Y, Müller CW (2018) The cryo-EM resolution revolution and transcription complexes. *Curr Opin Struct Biol* 52:8–15. <https://doi.org/10.1016/j.sbi.2018.07.002>
- Hao Y, Yu X, Bai Y, McBride HJ, Huang X (2019) Cryo-EM Structure of HER2-trastuzumab-pertuzumab complex. *Plos One* 14(5):e0216095. <https://doi.org/10.1371/journal.pone.0216095>

- Harastani M, Sorzano COS, Jonić S (2020) Hybrid Electron Microscopy Normal Mode Analysis with Scipion. *Protein Sci* 29(1):223–236. <https://doi.org/10.1002/pro.3772>
- Harastani M, Vuillemot R, Hamitouche I, Moghadam NB, Jonic S (2022) ContinuousFlex: Software package for analyzing continuous conformational variability of macromolecules in cryo electron microscopy and tomography data. *J Struct Biol* 214(4):107906. <https://doi.org/10.1016/j.jsb.2022.107906>
- HAS (2016) HAS-Herceptin(trastuzumab)-Gastric Cancer. [https://www.has-sante.fr/jcms/c\\_2633563/en/herceptin-trastuzumab-monoclonal-antibody-gastric-cancer](https://www.has-sante.fr/jcms/c_2633563/en/herceptin-trastuzumab-monoclonal-antibody-gastric-cancer). Accessed 1 Sep 2023
- Howie LJ, Scher NS, Amiri-Kordestani L, Zhang L, King-Kallimanis BL, Choudhry Y, Schroeder J, Goldberg KB, Kluetz PG, Ibrahim A, Sridhara R, Blumenthal GM, Pazdur R, Beaver JA (2019) FDA Approval Summary: Pertuzumab for Adjuvant Treatment of HER2-Positive Early Breast Cancer. *Clin Cancer Res* 25(10):2949–2955. <https://doi.org/10.1158/1078-0432.ccr-18-3003>
- Hung C-H, Yen J-B, Chang P-J, Chen L-W, Huang T-Y, Tsai W-J, Tsai Y-C (2023) Characterization of Human Norovirus Nonstructural Protein NS1.2 Involved in the Induction of the Filamentous Endoplasmic Reticulum, Enlarged Lipid Droplets, LC3 Recruitment, and Interaction with NTPase and NS4. *Viruses* 15(3):812. <https://doi.org/10.3390/v15030812>
- Huntington B, Zhao L, Bron P, Hameed UFS, Arold ST, Qureshi BM (2022) Thicker Ice Improves the Integrity and Angular Distribution of CDC48A Hexamers on Cryo-EM Grids. *Front Mol Biosci* 9:890390. <https://doi.org/10.3389/fmolb.2022.890390>
- Hutchison JM, Lu Z, Li GC, Travis B, Mittal R, Deatherage CL, Sanders CR (2017) Dodecyl- $\beta$ -melibioside Detergent Micelles as a Medium for Membrane Proteins. *Biochemistry* 56(41):5481–5484. <https://doi.org/10.1021/acs.biochem.7b00810>
- Iltchev A, Jensen J, Ellis D (2019) Electron Sources. <https://www.thermofisher.com/blog/materials/electron-source-fundamentals/>. Accessed 13 Sep 2023
- Jaiswal BS, Kljavin NM, Stawiski EW, Chan E, Parikh C, Durinck S, Chaudhuri S, Pujara K, Guillory J, Edgar KA, Janakiraman V, Scholz R-P, Bowman KK, Lorenzo M, Li H, Wu J, Yuan W, Peters BA, Kan Z, Stinson J, Mak M, Modrusan Z, Eigenbrot C, Firestein R, Stern HM, Rajalingam K, Schaefer G, Merchant MA, Sliwkowski MX, de Sauvage FJ, Seshagiri S (2013) Oncogenic ERBB3 Mutations in Human Cancers. *Cancer Cell* 23(5):603–617. <https://doi.org/10.1016/j.ccr.2013.04.012>
- Janeway CA, P T, Walport M, Shlomchik MJ (2001) The generation of diversity in immunoglobulins - Immunobiology, 5th edition. <https://www.ncbi.nlm.nih.gov/books/NBK27140/>. Accessed 2 Sep 2023
- JEOL (2023) JEOL Ltd. wavelength of electron. <https://www.jeol.com/words/emterms/20121023.071258.php#gsc.tab=0>. Accessed 13 Sep 2023



- Jerabek-Willemsen M, André T, Wanner R, Roth HM, Duhr S, Baaske P, Breitsprecher D (2014) MicroScale Thermophoresis: Interaction analysis and beyond. *J Mol Struct* 1077:101–113. <https://doi.org/10.1016/j.molstruc.2014.03.009>
- Kampjut D, Steiner J, Sazanov LA (2021) Cryo-EM grid optimization for membrane proteins. *Iscience* 24(3):102139. <https://doi.org/10.1016/j.isci.2021.102139>
- Karunagaran D, Tzahar E, Beerli RR, Chen X, Graus-Porta D, Ratzkin BJ, Seger R, Hynes NE, Yarden Y (1996) ErbB-2 is a common auxiliary subunit of NDF and EGF receptors: implications for breast cancer. *Embo J* 15(2):254–64
- Kato R, Naya M, Kasahata N, Senga R, Sato C, Koshino M, Suenaga K, Hasegawa M (2021) Thermal management function of graphene under cryogenic temperature. *Carbon* 183:970–976. <https://doi.org/10.1016/j.carbon.2021.07.077>
- Klebl DP, Gravett MSC, Kontziampasis D, Wright DJ, Bon RS, Monteiro DCF, Trebbin M, Sobott F, White HD, Darrow MC, Thompson RF, Muench SP (2020) Need for Speed: Examining Protein Behavior during CryoEM Grid Preparation at Different Timescales. *Structure* 28(11):1238-1248.e4. <https://doi.org/10.1016/j.str.2020.07.018>
- Kobayashi C, Jung J, Matsunaga Y, Mori T, Ando T, Tamura K, Kamiya M, Sugita Y (2017) GENESIS 1.1: A hybrid-parallel molecular dynamics simulator with enhanced sampling algorithms on multiple computational platforms. *J Comput Chem* 38(25):2193–2206. <https://doi.org/10.1002/jcc.24874>
- KÖHLER G, MILSTEIN C (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 256(5517):495–497. <https://doi.org/10.1038/256495a0>
- Kojima Y, Sudo K, Yoshida H, Yazaki S, Tokura M, Mizoguchi C, Okuma HS, Kita S, Yamamoto K, Nishikawa T, Noguchi E, Shimoi T, Tanase Y, Uno M, Ishikawa M, Kato T, Koyama K, Kobayashi M, Kakegawa T, Fujiwara Y, Yonemori K (2023) Changes in HER3 expression profiles between primary and recurrent gynecological cancers. *Cancer Cell Int* 23(1):18. <https://doi.org/10.1186/s12935-022-02844-z>
- Kühlbrandt W (2014) The Resolution Revolution. *Science* 343(6178):1443–1444. <https://doi.org/10.1126/science.1251652>
- Ledsgaard L, Kilstrup M, Karatt-Vellatt A, McCafferty J, Laustsen AH (2018) Basics of Antibody Phage Display Technology. *Toxins* 10(6):236. <https://doi.org/10.3390/toxins10060236>
- Leyton JV (2020) Improving Receptor-Mediated Intracellular Access and Accumulation of Antibody Therapeutics—The Tale of HER2. *Antibodies* 9(3):32. <https://doi.org/10.3390/antib9030032>
- Lipfert J, Columbus L, Chu VB, Lesley SA, Doniach S (2007) Size and Shape of Detergent Micelles Determined by Small-Angle X-ray Scattering. *J Phys Chem B* 111(43):12427–12438. <https://doi.org/10.1021/jp073016l>

- Little AJ, Matthews A, Oettinger M, Roth DB, Schatz DG (2015) Molecular Biology of B Cells (Second Edition). :13–34. <https://doi.org/10.1016/b978-0-12-397933-9.00002-3>
- Liu X, Fang Y, Li Y, Li Y, Qi L, Wang X (2022) Pertuzumab combined with trastuzumab compared to trastuzumab in the treatment of HER2-positive breast cancer: A systematic review and meta-analysis of randomized controlled trials. *Frontiers Oncol* 12:894861. <https://doi.org/10.3389/fonc.2022.894861>
- Lucas LM, Dwivedi V, Senfeld JI, Cullum RL, Mill CP, Piazza JT, Bryant IN, Cook LJ, Miller ST, Lott JH, Kelley CM, Knerr EL, Markham JA, Kaufmann DP, Jacobi MA, Shen J, Riese DJ (2022) The Yin and Yang of ERBB4: Tumor Suppressor and Oncoprotein. *Pharmacol Rev* 74(1):18–47. <https://doi.org/10.1124/pharmrev.121.000381>
- Mariani V, Biasini M, Barbato A, Schwede T (2013) IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* 29(21):2722–2728. <https://doi.org/10.1093/bioinformatics/btt473>
- Marintcheva B (2018) Harnessing the Power of Viruses. (*Proc Natl Acad Sci USA* 106 5 2009):133–160. <https://doi.org/10.1016/b978-0-12-810514-6.00005-2>
- Mayer G, Hudrisier D (2012) Immunology 7ème édition - GENETIQUE DES IMMUNOGLOBULINES. <https://www.microbiologybook.org/French-immuno/immchapter6.htm>. Accessed 2 Sep 2023
- Medina KL (2016) Chapter 4 Overview of the immune system. *Handb Clin Neurol* 133:61–76. <https://doi.org/10.1016/b978-0-444-63432-0.00004-9>
- Molina MA, Codony-Servat J, Albanell J, Rojo F, Arribas J, Baselga J (2001) Trastuzumab (herceptin), a humanized anti-Her2 receptor monoclonal antibody, inhibits basal and activated Her2 ectodomain cleavage in breast cancer cells. *Cancer Res* 61(12):4744–9
- Mosavi LK, Minor DL, Peng Z (2002) Consensus-derived structural determinants of the ankyrin repeat motif. *Proc National Acad Sci* 99(25):16029–16034. <https://doi.org/10.1073/pnas.252537899>
- Mukherjee S, Erramilli SK, Ammirati M, Alvarez FJD, Fennell KF, Purdy MD, Skrobek BM, Radziwon K, Coukos J, Kang Y, Dutka P, Gao X, Qiu X, Yeager M, Xu HE, Han S, Kossiakoff AA (2020) Synthetic antibodies against BRIL as universal fiducial marks for single-particle cryoEM structure determination of membrane proteins. *Nat Commun* 11(1):1598. <https://doi.org/10.1038/s41467-020-15363-0>
- Nahta R, Hung M-C, Esteva FJ (2004) The HER-2-Targeting Antibodies Trastuzumab and Pertuzumab Synergistically Inhibit the Survival of Breast Cancer Cells. *Cancer Res* 64(7):2343–2346. <https://doi.org/10.1158/0008-5472.can-03-3856>
- Nami B, Maadi H, Wang Z (2018) Mechanisms Underlying the Action and Synergism of Trastuzumab and Pertuzumab in Targeting HER2-Positive Breast Cancer. *Cancers* 10(10):342. <https://doi.org/10.3390/cancers10100342>

- Nangola S, Urvoas A, Valerio-Lepiniec M, Khamaikawin W, Sakkhachornphop S, Hong S-S, Boulanger P, Minard P, Tayapiwatana C (2012) Antiviral activity of recombinant ankyrin targeted to the capsid domain of HIV-1 Gag polyprotein. *Retrovirology* 9(1):17. <https://doi.org/10.1186/1742-4690-9-17>
- Naydenova K, Peet MJ, Russo CJ (2019) Multifunctional graphene supports for electron cryomicroscopy. *Proc Natl Acad Sci* 116(24):11718–11724. <https://doi.org/10.1073/pnas.1904766116>
- Neselu K, Wang B, Rice WJ, Potter CS, Carragher B, Chua EYD (2023) Measuring the effects of ice thickness on resolution in single particle cryo-EM. *J Struct Biol: X* 7:100085. <https://doi.org/10.1016/j.yjsbx.2023.100085>
- Nogales E, Scheres SHW (2015) Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Mol Cell* 58(4):677–689. <https://doi.org/10.1016/j.molcel.2015.02.019>
- Olayioye MA, Neve RM, Lane HA, Hynes NE (2000) The ErbB signaling network: receptor heterodimerization in development and cancer. *Embo J* 19(13):3159–3167. <https://doi.org/10.1093/emboj/19.13.3159>
- Passaro A, Jänne PA, Mok T, Peters S (2021) Overcoming therapy resistance in EGFR-mutant lung cancer. *Nat Cancer* 2(4):377–391. <https://doi.org/10.1038/s43018-021-00195-8>
- Passmore LA, Russo CJ (2016) Chapter Three Specimen Preparation for High-Resolution Cryo-EM. *Methods Enzymol* 579:51–86. <https://doi.org/10.1016/bs.mie.2016.04.011>
- Peckys DB, Gaa D, Jonge N de (2021) Quantification of EGFR-HER2 Heterodimers in HER2-Overexpressing Breast Cancer Cells Using Liquid-Phase Electron Microscopy. *Cells* 10(11):3244. <https://doi.org/10.3390/cells10113244>
- Peckys DB, Korf U, Jonge N de (2015) Local variations of HER2 dimerization in breast cancer cells discovered by correlative fluorescence and liquid electron microscopy. *Sci Adv* 1(6):e1500165. <https://doi.org/10.1126/sciadv.1500165>
- Penczek PA (2010) Chapter Three Resolution Measures in Molecular Electron Microscopy. *Methods Enzym* 482:73–100. [https://doi.org/10.1016/s0076-6879\(10\)82003-8](https://doi.org/10.1016/s0076-6879(10)82003-8)
- Pohlmann PR, Mayer IA, Mernaugh R (2009) Resistance to Trastuzumab in Breast Cancer. *Clin Cancer Res* 15(24):7479–7491. <https://doi.org/10.1158/1078-0432.ccr-09-0636>
- Provencher SW (1982) A constrained regularization method for inverting data represented by linear algebraic or integral equations. *Comput Phys Commun* 27(3):213–227. [https://doi.org/10.1016/0010-4655\(82\)90173-4](https://doi.org/10.1016/0010-4655(82)90173-4)
- Punjani A, Fleet DJ (2021) 3D Flexible Refinement: Structure and Motion of Flexible Proteins from Cryo-EM. *Biorxiv* :2021.04.22.440893. <https://doi.org/10.1101/2021.04.22.440893>

- Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* 14(3):290–296. <https://doi.org/10.1038/nmeth.4169>
- Rees AR (2020) Understanding the human antibody repertoire. *mAbs* 12(1):1729683. <https://doi.org/10.1080/19420862.2020.1729683>
- Ren X-R, Wei J, Lei G, Wang J, Lu J, Xia W, Spector N, Barak LS, Clay TM, Osada T, Hamilton E, Blackwell K, Hobeika AC, Morse MA, Lyerly HK, Chen W (2012) Polyclonal HER2-specific antibodies induced by vaccination mediate receptor internalization and degradation in tumor cells. *Breast Cancer Res* 14(3):R89. <https://doi.org/10.1186/bcr3204>
- Rice WJ, Cheng A, Noble AJ, Eng ET, Kim LY, Carragher B, Potter CS (2018) Routine determination of ice thickness for cryo-EM grids. *J Struct Biol* 204(1):38–44. <https://doi.org/10.1016/j.jsb.2018.06.007>
- Richard S, Selle F, Lotz J-P, Khalil A, Gligorov J, Soares DG (2016) Pertuzumab and trastuzumab: the rationale way to synergy. *An da Acad Bras Ciências* 88(Suppl 1):565–577. <https://doi.org/10.1590/0001-3765201620150178>
- Robinson PJ, Bulleid NJ (2020) Mechanisms of Disulfide Bond Formation in Nascent Polypeptides Entering the Secretory Pathway. *Cells* 9(9):1994. <https://doi.org/10.3390/cells9091994>
- Rosa-Trevín JM de la, Quintana A, Cano L del, Zaldívar A, Foche I, Gutiérrez J, Gómez-Blanco J, Burguet-Castell J, Cuenca-Alba J, Abrishami V, Vargas J, Otón J, Sharov G, Vilas JL, Navas J, Conesa P, Kazemi M, Marabini R, Sorzano COS, Carazo JM (2016) Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J Struct Biol* 195(1):93–99. <https://doi.org/10.1016/j.jsb.2016.04.010>
- Rose DR (1982) The generation of antibody diversity. *Am J Hematol* 13(1):91–99. <https://doi.org/10.1002/ajh.2830130111>
- Rosenthal PB, Henderson R (2003) Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *J Mol Biol* 333(4):721–745. <https://doi.org/10.1016/j.jmb.2003.07.013>
- Roth DB (2014) V(D)J Recombination: Mechanism, Errors, and Fidelity. *Microbiol Spectr* 2(6). <https://doi.org/10.1128/microbiolspec.mdna3-0041-2014>
- Russell SD, university O, Larson P (2019) Accelerating Voltage Physics Calculator. <https://www.ou.edu/research/electron/bmz5364/calc-kv.html>. Accessed 13 Sep 2023
- Russo CJ, Passmore LA (2016) Ultrastable gold substrates: Properties of a support for high-resolution electron cryomicroscopy of biological specimens. *J Struct Biol* 193(1):33–44. <https://doi.org/10.1016/j.jsb.2015.11.006>

- Safdari Y, Farajnia S, Asgharzadeh M, Khalili M (2013) Antibody humanization methods – a review and update. *Biotechnol Genet Eng Rev* 29(2):175–186. <https://doi.org/10.1080/02648725.2013.801235>
- Sakkhachornphop S, Hadpech S, Wisitponchai T, Panto C, Kantamala D, Utaipat U, Preparattanapan J, Kotarathitithum W, Taejaroenkul S, Yasamut U, Chupradit K, Moonmuang S, Lee VS, Suparatpinyo K, Tayapiwatana C (2018) Broad-Spectrum Antiviral Activity of an Ankyrin Repeat Protein on Viral Assembly against Chimeric NL4-3 Viruses Carrying Gag/PR Derived from Circulating Strains among Northern Thai Patients. *Viruses* 10(11):625. <https://doi.org/10.3390/v10110625>
- Scheres SHW (2012) RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* 180(3):519–530. <https://doi.org/10.1016/j.jsb.2012.09.006>
- Scheres SHW, Chen S (2012) Prevention of overfitting in cryo-EM structure determination. *Nat Methods* 9(9):853–854. <https://doi.org/10.1038/nmeth.2115>
- Scheuer W, Friess T, Burtscher H, Bossenmaier B, Endl J, Hasmann M (2009) Strongly Enhanced Antitumor Activity of Trastuzumab and Pertuzumab Combination Treatment on HER2-Positive Human Xenograft Tumor Models. *Cancer Res* 69(24):9330–9336. <https://doi.org/10.1158/0008-5472.can-08-4597>
- Schroeder HW, Cavacini L (2010) Structure and function of immunoglobulins. *J Allergy Clin Immunol* 125(2):S41–S52. <https://doi.org/10.1016/j.jaci.2009.09.046>
- Scotti A, Liu W, Hyatt JS, Herman ES, Choi HS, Kim JW, Lyon LA, Gasser U, Fernandez-Nieves A (2015) The CONTIN algorithm and its application to determine the size distribution of microgel suspensions. *J Chem Phys* 142(23):234905. <https://doi.org/10.1063/1.4921686>
- Stumpff MT, Binz HK, Amstutz P (2008) DARPins: A new generation of protein therapeutics. *Drug Discov Today* 13(15–16):695–701. <https://doi.org/10.1016/j.drudis.2008.04.013>
- Stumpff MT, Dawson KM, Binz HK (2020) Beyond Antibodies: The DARPin® Drug Platform. *BioDrugs* 34(4):423–433. <https://doi.org/10.1007/s40259-020-00429-8>
- Teterina NL, Pinto Y, Weaver JD, Jensen KS, Ehrenfeld E (2011) Analysis of Poliovirus Protein 3A Interactions with Viral and Cellular Proteins in Infected Cells. *J Virol* 85(9):4284–4296. <https://doi.org/10.1128/jvi.02398-10>
- Tran TNH, Trainor E, Nakagomi T, Cunliffe NA, Nakagomi O (2013) Molecular epidemiology of noroviruses associated with acute sporadic gastroenteritis in children: Global distribution of genogroups, genotypes and GI.4 variants. *J Clin Virol* 56(3):269–277. <https://doi.org/10.1016/j.jcv.2012.11.011>
- Tsumoto K, Umetsu M, Kumagai I, Ejima D, Philo JS, Arakawa T (2004) Role of Arginine in Protein Refolding, Solubilization, and Purification. *Biotechnol Progr* 20(5):1301–1308. <https://doi.org/10.1021/bp0498793>

- Uchański T, Masiulis S, Fischer B, Kalichuk V, López-Sánchez U, Zarkadas E, Weckener M, Sente A, Ward P, Wohlkönig A, Zögg T, Remaut H, Naismith JH, Nury H, Vranken W, Aricescu AR, Pardon E, Steyaert J (2021) Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. *Nat Methods* 18(1):60–68. <https://doi.org/10.1038/s41592-020-01001-6>
- Urvoas A, Guellouz A, Valerio-Lepiniec M, Graille M, Durand D, Desravines DC, Tilbeurgh H van, Desmadril M, Minard P (2010) Design, Production and Molecular Structure of a New Family of Artificial Alpha-helicoidal Repeat Proteins ( $\alpha$ Rep) Based on Thermostable HEAT-like Repeats. *J Mol Biol* 404(2):307–327. <https://doi.org/10.1016/j.jmb.2010.09.048>
- Urvoas A, Valerio-Lepiniec M, Minard P (2012) Artificial proteins from combinatorial approaches. *Trends Biotechnol* 30(10):512–520. <https://doi.org/10.1016/j.tibtech.2012.06.001>
- Valerio-Lepiniec M, Urvoas A, Chevrel A, Guellouz A, Ferrandez Y, Mesneau A, Sierra-Gallay IL de la, Aumont-Nicaise M, Desmadril M, Tilbeurgh H van, Minard P (2015) The  $\alpha$ Rep artificial repeat protein scaffold: a new tool for crystallization and live cell applications. *Biochem Soc Trans* 43(5):819–824. <https://doi.org/10.1042/bst20150075>
- Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Židek A, Green T, Tunyasuvunakool K, Petersen S, Jumper J, Clancy E, Green R, Vora A, Lutfi M, Figurnov M, Cowie A, Hobbs N, Kohli P, Kleywegt G, Birney E, Hassabis D, Velankar S (2021) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 50(D1):D439–D444. <https://doi.org/10.1093/nar/gkab1061>
- Vinothkumar KR, Henderson R (2016) Single particle electron cryomicroscopy: trends, issues and future perspective. *Q Rev Biophys* 49:e13. <https://doi.org/10.1017/s0033583516000068>
- Vuillemot R, Mirzaei A, Harastani M, Hamitouche I, Fréchin L, Klaholz BP, Miyashita O, Tama F, Rouiller I, Jonic S (2023) MDSPACE: Extracting continuous conformational landscapes from cryo-EM single particle datasets using 3D-to-2D flexible fitting based on Molecular Dynamics simulation. *J Mol Biol* :167951. <https://doi.org/10.1016/j.jmb.2023.167951>
- Vuillemot R, Miyashita O, Tama F, Rouiller I, Jonic S (2022) NMMD: Efficient Cryo-EM Flexible Fitting Based on Simultaneous Normal Mode and Molecular Dynamics atomic displacements. *J Mol Biol* 434(7):167483. <https://doi.org/10.1016/j.jmb.2022.167483>
- Wagner T, Merino F, Stabrin M, Moriya T, Antoni C, Apelbaum A, Hagel P, Sitsel O, Raisch T, Prumbaum D, Quentin D, Roderer D, Tacke S, Siebolds B, Schubert E, Shaikh TR, Lill P, Gatsogiannis C, Raunser S (2019) SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun Biol* 2(1):218. <https://doi.org/10.1038/s42003-019-0437-z>
- Walser M, Mayor J, Rothenberger S (2022) Designed Ankyrin Repeat Proteins: A New Class of Viral Entry Inhibitors. *Viruses* 14(10):2242. <https://doi.org/10.3390/v14102242>

- Wang J, Ptacek JB, Kirkegaard K, Bullitt E (2013) Double-membraned Liposomes Sculpted by Poliovirus 3AB Protein\*. *J Biol Chem* 288(38):27287–27298. <https://doi.org/10.1074/jbc.m113.498899>
- Weissenberger G, Henderikx RJM, Peters PJ (2021) Understanding the invisible hands of sample preparation for cryo-EM. *Nat Methods* 18(5):463–471. <https://doi.org/10.1038/s41592-021-01130-6>
- Wieduwilt MJ, Moasser MM (2008) The epidermal growth factor receptor family: Biology driving targeted therapeutics. *Cell Mol Life Sci* 65(10):1566–1584. <https://doi.org/10.1007/s00018-008-7440-8>
- Williams DB, Carter CB (2009) *Transmission Electron Microscopy, A Textbook for Materials Science*. <https://doi.org/10.1007/978-0-387-76501-3>
- Wu M, Lander GC (2020) How low can we go? Structure determination of small biological complexes using single-particle cryo-EM. *Curr Opin Struct Biol* 64:9–16. <https://doi.org/10.1016/j.sbi.2020.05.007>
- Wu S, Armache J-P, Cheng Y (2015) Single-particle cryo-EM data acquisition by using direct electron detection camera. *Microscopy* 65(1):35–41. <https://doi.org/10.1093/jmicro/dfv355>
- Young G, Hundt N, Cole D, Fineberg A, Andrecka J, Tyler A, Olerinyova A, Ansari A, Marklund EG, Collier MP, Chandler SA, Tkachenko O, Allen J, Crispin M, Billington N, Takagi Y, Sellers JR, Eichmann C, Selenko P, Frey L, Riek R, Galpin MR, Struwe WB, Benesch JLP, Kukura P (2018) Quantitative mass imaging of single biological macromolecules. *Science* 360(6387):423–427. <https://doi.org/10.1126/science.aar5839>
- Zhang K (2016) Gctf: Real-time CTF determination and correction. *J Struct Biol* 193(1):1–12. <https://doi.org/10.1016/j.jsb.2015.11.003>
- Zivanov J, Nakane T, Scheres SHW (2019) A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* 6(1):5–17. <https://doi.org/10.1107/s205225251801463x>

## 6 Annexes

### 6.1 Alignements des différent génogroupes de NS4

#### 6.1.1 Alignement des séquences de GI

	1	10	20	30	40	50	60						
p22_GI.3/Human/United	GKM	YDFD	DRVS	AFT	MARD	NCLG	IISMA	LGKKLR	GVT	MEGLKNAI	KCYKI	SAC	IKW
p22_Hu/GI.1/CHA9A004_20110419/2011/USA	GPT	YDFD	DRVA	AFT	MARD	NCLG	IISMA	LGKKLR	SVT	IEGLKNAI	SCYKI	SKC	SIQW
p22_Hu/GI.1/CHA6A003_20091031/2009/USA	GPT	YDFD	DRVA	AFT	MARD	NCLG	IISMA	LGKKLR	SVT	IEGLKNAI	SCYKI	SKC	SIQW
p22_Hu/GI.2/Jingzhou/2013401/CHN	GKM	YDFD	DRVS	AFT	MARD	NCLG	IISMA	LGKKLR	GVT	MEGLKNAI	KCYKI	SAC	IKW
p22_Hu/GI.2/Leuven/2003/BEL	GKM	YDFD	DRVS	AFT	MARD	NCLG	IISMA	LGKKLR	GVT	MEGLKNAI	KCYKI	SAC	IKW
p22_Hu/GI.4/Groningen/2011/NL	GKM	YDFD	DRVS	AFT	MARD	NCLG	IISMA	LGKKLR	GVT	MEGLKNAI	KCYKI	SAC	IKW
p22_Hu/GI.6/Groningen/2012/NL	GKM	YDFD	DRVS	AFT	MARD	NCLG	IISMA	LGKKLR	GVT	MEGLKNAI	KCYKI	SAC	IKW
Hu/GI.6/Kingston/ACT160D/2010/AU	GKM	YDFD	DRVS	AFT	MARD	NCLG	IISMA	LGKKLR	GVT	MEGLKNAI	KCYKI	SAC	IKW

	70	80	90	100	110	120					
p22_GI.3/Human/United	QARVYS	LESDGN	SVNIR	EERNIL	TQQ	QOSVCA	ASVAL	TRLR	AARAVAYASC	FSAIT	STIL
p22_Hu/GI.1/CHA9A004_20110419/2011/USA	QSRVYI	LESDG	SVQIR	EDKQAL	TPIQ	QOINT	ASAI	TRIK	AARAVAYASC	FSAIT	STIL
p22_Hu/GI.1/CHA6A003_20091031/2009/USA	QSRVYI	LESDG	SVQIR	EDKQAL	TPIQ	QOINT	ASAI	TRIK	AARAVAYASC	FSAIT	STIL
p22_Hu/GI.2/Jingzhou/2013401/CHN	QARVYS	LESDGN	SVNIR	EERNIL	TQQ	QOSVCA	ASVAL	TRLR	AARAVAYASC	FSAIT	STIL
p22_Hu/GI.2/Leuven/2003/BEL	QARVYS	LESDGN	SVNIR	EERNIL	TQQ	QOSVCA	ASVAL	TRLR	AARAVAYASC	FSAIT	STIL
p22_Hu/GI.4/Groningen/2011/NL	QARVYS	LESDGN	SVNIR	EERNIL	TQQ	QOSVCA	ASVAL	TRLR	AARAVAYASC	FSAIT	STIL
p22_Hu/GI.6/Groningen/2012/NL	QARVYS	LESDGN	SVNIR	EERNIL	TQQ	QOSVCA	ASVAL	TRLR	AARAVAYASC	FSAIT	STIL
Hu/GI.6/Kingston/ACT160D/2010/AU	QARVYS	LESDGN	SVNIR	EERNIL	TQQ	QOSVCA	ASVAL	TRLR	AARAVAYASC	FSAIT	STIL

	130	140	150	160	170	180								
p22_GI.3/Human/United	QIAC	SALV	NRAV	KRMFG	TRTA	TLS	LEGPP	EEHR	CRVH	MAKA	ACKGPI	GHDD	WVEKY	GLC
p22_Hu/GI.1/CHA9A004_20110419/2011/USA	QIAC	SALV	NRAV	KRMFG	TRTA	AMA	LEGPP	GHEHN	CRVB	RAKE	ACKGPI	GHDD	MVERE	GLC
p22_Hu/GI.1/CHA6A003_20091031/2009/USA	QIAC	SALV	NRAV	KRMFG	TRTA	AMA	LEGPP	GHEHN	CRVB	RAKE	ACKGPI	GHDD	MVERE	GLC
p22_Hu/GI.2/Jingzhou/2013401/CHN	QIAC	SALV	NRAV	KRMFG	TRTA	TLS	LEGPP	EEHR	CRVH	MAKA	ACKGPI	GHDD	WVEKY	GLC
p22_Hu/GI.2/Leuven/2003/BEL	QIAC	SALV	NRAV	KRMFG	TRTA	TLS	LEGPP	EEHR	CRVH	MAKA	ACKGPI	GHDD	WVEKY	GLC
p22_Hu/GI.4/Groningen/2011/NL	QIAC	SALV	NRAV	KRMFG	TRTA	TLS	LEGPP	EEHR	CRVH	MAKA	ACKGPI	GHDD	WVEKY	GLC
p22_Hu/GI.6/Groningen/2012/NL	QIAC	SALV	NRAV	KRMFG	TRTA	TLS	LEGPP	EEHR	CRVH	MAKA	ACKGPI	GHDD	WVEKY	GLC
Hu/GI.6/Kingston/ACT160D/2010/AU	QIAC	SALV	NRAV	KRMFG	TRTA	TLS	LEGPP	EEHR	CRVH	MAKA	ACKGPI	GHDD	WVEKY	GLC

	190			
p22_GI.3/Human/United	ETEDDE	EVAHT	EIPSAT	ME..
p22_Hu/GI.1/CHA9A004_20110419/2011/USA	ETEDDE	SEDOI	QMVP	SDAVPE
p22_Hu/GI.1/CHA6A003_20091031/2009/USA	ETEDDE	SEDOI	QMVP	SDAVPE
p22_Hu/GI.2/Jingzhou/2013401/CHN	ETEDDE	EVAHT	EIPSAT	ME..
p22_Hu/GI.2/Leuven/2003/BEL	ETEDDE	EVAHA	EIPSAT	ME..
p22_Hu/GI.4/Groningen/2011/NL	ETEDDE	EVAHT	EIPSAT	ME..
p22_Hu/GI.6/Groningen/2012/NL	ETEDDE	EVVHT	EMPSAT	IE..
Hu/GI.6/Kingston/ACT160D/2010/AU	ETEDDE	EVAQT	EMPSAT	IE..





## 6.1.4 Alignement des différentes séquences des différents génogroupes

	1	10	20	30	40	50	60																																																						
p22_Hu/GI.1/CHA9A004_20110419/2011/USA	G	P	T	Y	D	F	D	T	D	R	V	A	A	F	T	R	M	A	R	A	N	G	L	G	L	I	S	M	A	S	L	G	K	K	L	R	S	V	T	T	E	G	T	K	N	A	L	S	G	Y	R	I	S	K	C	S	I	Q	W		
p22_Hu/GI.2/Leuven/2003/BEL	G	K	M	Y	D	F	D	D	D	R	V	S	A	F	T	T	M	A	R	D	N	G	L	G	I	L	S	M	A	G	L	G	K	K	L	R	G	V	T	T	M	E	G	L	K	N	A	L	K	G	Y	K	I	S	A	C	T	I	K	W	
p22_GI.3/Human/United	G	K	M	Y	D	F	D	D	D	R	V	S	A	F	T	T	M	A	R	D	N	G	L	G	I	L	S	M	A	G	L	G	K	K	L	R	G	V	T	T	M	E	G	L	K	N	A	L	K	G	Y	K	I	S	A	C	T	I	K	W	
p22_Hu/GI.4/Groningen/2011/NL	G	K	M	Y	D	F	D	D	D	R	V	S	A	F	T	T	M	A	R	D	N	G	L	G	I	L	S	M	A	G	L	G	K	K	L	R	G	V	T	T	M	E	G	L	K	N	A	L	K	G	Y	K	I	S	A	C	T	I	K	W	
p22_Hu/GI.6/Groningen/2012/NL	G	K	M	Y	D	F	D	D	D	R	V	S	A	F	T	T	M	A	R	D	N	G	L	G	I	L	S	M	A	S	L	G	K	K	L	R	G	V	T	S	M	E	G	L	K	N	A	L	K	G	Y	K	I	G	A	C	T	I	K	W	
NS4_Hu/GII.4/Helensburgh/NSW295E/2010/AU	C	P	A	L	T	T	F	N	F	D	R	N	K	V	L	A	F	R	Q	L	A	A	E	N	K	Y	G	L	M	D	T	M	R	V	G	K	L	K	D	V	R	T	M	P	E	L	R	Q	A	L	K	N	V	S	I	K	K	C	Q		
NS4_Hu/GII.4/Berowra/NSW767L/2012/AU	C	P	A	L	T	T	F	N	F	D	R	N	K	V	L	A	F	R	Q	L	A	A	E	N	K	Y	G	L	M	D	T	M	R	V	G	K	L	K	D	V	R	T	M	P	E	L	R	Q	A	L	K	N	V	S	I	K	K	C	Q		
NS4_Hu/GII.3	C	P	T	P	T	T	F	N	F	D	N	K	V	F	A	F	R	Q	L	A	A	E	N	K	Y	G	L	M	D	T	M	R	V	G	S	L	K	G	V	K	T	V	S	E	L	R	Q	A	L	K	N	I	A	I	R	C	Q				
NS4_Hu/GII.12/CGMH40/2010/TW	C	P	T	P	T	T	F	N	F	D	N	K	V	F	A	F	R	Q	L	A	A	E	N	K	Y	G	L	M	D	T	M	R	V	G	S	L	K	N	V	K	T	V	S	E	L	R	Q	A	L	K	N	I	A	I	R	C	Q				
NS4_Hu/GII.5/Saitama/T49/2001/JP	C	P	E	L	P	V	Y	N	F	D	H	N	R	V	A	A	F	R	L	A	A	D	N	K	Y	G	L	M	E	T	I	K	V	G	N	K	L	G	V	K	T	L	E	Q	L	K	D	A	L	K	D	I	S	I	M	P	C	R			
NS4_GIV.2	G	K	E	D	P	T	Y	N	F	D	S	L	P	V	T	A	F	R	R	L	A	A	D	N	G	Y	N	L	A	K	T	M	A	V	G	A	K	L	R	K	V	T	T	V	E	G	F	R	D	A	V	K	D	V	A	F	K	P	C	R	
		70	80	90	100	110	120																																																						
p22_Hu/GI.1/CHA9A004_20110419/2011/USA	Q	S	R	V	Y	I	E	S	D	G	A	S	V	Q	I	K	E	D	K	Q	A	L	T	P	L	Q	T	I	N	T	A	S	L	A	I	T	R	L	K	A	A	R	A	V	A	S	C	F	Q	S	A	I	T	T	I	D					
p22_Hu/GI.2/Leuven/2003/BEL	Q	A	K	V	S	L	E	S	D	G	N	S	V	N	I	K	E	E	R	N	I	L	T	Q	Q	Q	S	V	C	A	A	S	V	A	L	T	R	L	R	A	A	R	A	V	A	S	C	I	Q	S	A	I	T	S	I	L					
p22_GI.3/Human/United	Q	A	K	V	S	L	E	S	D	G	N	S	V	N	I	K	E	E	R	N	I	L	T	Q	Q	Q	S	V	C	A	A	S	V	A	L	T	R	L	R	A	A	R	A	V	A	S	C	I	Q	S	A	I	T	S	I	L					
p22_Hu/GI.4/Groningen/2011/NL	Q	A	K	V	S	L	E	S	D	G	N	S	V	N	I	K	E	E	R	N	I	L	T	Q	Q	Q	S	V	C	A	A	S	V	A	L	T	R	L	R	A	A	R	A	V	A	S	C	T	Q	S	A	I	T	S	I	L					
p22_Hu/GI.6/Groningen/2012/NL	Q	A	K	V	S	L	E	S	D	G	N	S	V	N	I	K	E	E	R	N	I	L	T	Q	Q	Q	S	V	C	A	A	S	I	A	L	T	R	L	R	A	A	R	A	V	A	S	C	I	Q	S	A	I	T	S	I	L					
NS4_Hu/GII.4/Helensburgh/NSW295E/2010/AU	I	V	Y	S	G	C	T	Y	L	E	S	D	G	K	N	V	K	V	D	R	I	Q	S	A	A	V	Q	T	N	N	E	L	A	G	A	L	H	L	R	C	A	R	I	R	Y	V	K	C	V	Q	E	A	L	Y	S	I					
NS4_Hu/GII.4/Berowra/NSW767L/2012/AU	I	V	Y	S	G	C	T	Y	L	E	S	D	G	K	N	V	K	V	D	R	I	Q	S	A	A	V	Q	T	N	N	E	L	A	G	A	L	H	L	R	C	A	R	I	R	Y	V	K	C	V	Q	E	A	L	Y	S	I					
NS4_Hu/GII.3	I	V	Y	S	G	C	T	Y	L	E	S	D	G	K	N	V	K	V	D	R	I	Q	S	A	A	V	Q	T	N	N	E	L	S	G	A	L	H	L	R	C	A	R	I	R	Y	V	K	C	V	Q	E	A	L	Y	S	I					
NS4_Hu/GII.12/CGMH40/2010/TW	I	V	Y	S	G	L	T	Y	S	L	E	S	D	G	K	D	V	R	V	E	K	V	Q	S	P	A	V	Q	T	N	N	E	L	T	G	A	L	H	L	R	C	A	R	I	R	Y	V	K	C	V	Q	E	A	L	Y	S	I				
NS4_Hu/GII.5/Saitama/T49/2001/JP	I	V	Y	N	S	M	A	Y	D	L	S	N	G	S	G	R	V	S	V	K	K	V	V	D	H	K	I	Q	T	N	N	E	L	N	S	A	L	N	L	R	G	A	R	V	R	Y	V	K	C	V	Q	E	M	V	Y	S	L				
NS4_GIV.2	I	V	Y	K	G	V	A	Y	D	M	S	D	G	R	G	K	V	K	V	D	R	V	Q	D	Q	V	V	R	Q	T	E	V	H	Q	V	I	H	R	L	R	Q	A	R	V	R	Y	V	H	V	A	Q	E	L	V	M	S	L				
		130	140	150	160	170	180																																																						
p22_Hu/GI.1/CHA9A004_20110419/2011/USA	Q	M	A	G	S	A	L	V	V	N	R	A	V	K	R	M	F	G	T	R	A	A	M	A	L	E	G	P	G	K	E	H	N	C	R	V	H	K	A	K	E	A	G	K	G	P	I	G	H	D	D	M	V	E	R	F	G	L	C		
p22_Hu/GI.2/Leuven/2003/BEL	Q	I	A	G	S	A	L	V	V	N	R	A	V	K	R	M	F	G	T	R	A	T	L	S	L	E	G	P	P	R	E	H	K	C	R	V	H	M	A	K	A	A	A	G	K	G	P	I	G	H	D	D	V	V	E	K	Y	G	L	C	
p22_GI.3/Human/United	Q	I	A	G	S	A	L	V	V	N	R	A	V	K	R	M	F	G	T	R	A	T	L	S	L	E	G	P	P	R	E	H	K	C	R	V	H	M	A	K	A	A	A	A	G	K	G	P	I	G	H	D	D	V	V	E	K	Y	G	L	C
p22_Hu/GI.4/Groningen/2011/NL	Q	I	A	G	S	A	L	V	V	N	R	A	V	K	R	M	F	G	T	R	A	T	L	S	L	E	G	P	P	R	E	H	K	C	R	V	H	M	A	K	A	A	A	A	G	K	G	P	I	G	H	D	D	V	V	E	K	Y	G	L	C
p22_Hu/GI.6/Groningen/2012/NL	Q	I	A	S	A	L	V	V	N	R	A	V	K	R	M	F	G	T	R	A	A	L	S	L	E	G	P	P	K	E	H	K	C	R	V	H	Q	A	K	A	A	G	K	G	P	I	G	H	D	D	V	V	E	K	Y	G	L	C			
NS4_Hu/GII.4/Helensburgh/NSW295E/2010/AU	I	Q	I	A	G	A	A	F	V	T	T	R	I	A	K	R	M	N	I	Q	D	L	W	S	K	P	Q	V	E	N	T	E	E	A	T	S	K	D	G	C	P	K	P	K	D	E	E	L	V	I	S	S	D	D	I	K	T	E			
NS4_Hu/GII.4/Berowra/NSW767L/2012/AU	I	Q	I	A	G	A	A	F	V	T	T	R	I	A	K	R	M	N	I	Q	D	L	W	S	K	P	Q	V	E	N	T	E	E	T	S	K	D	G	C	P	K	P	K	D	E	E	F	V	I	S	S	D	D	I	K	T	E				
NS4_Hu/GII.3	I	Q	I	A	G	A	A	F	V	T	T	R	I	A	K	R	M	N	I	Q	N	L	W	S	R	P	Q	V	E	D	A	E	V	I	N	R	D	G	C	P	K	P	K	D	E	E	F	V	I	S	S	D	D	I	K	A	E				
NS4_Hu/GII.12/CGMH40/2010/TW	I	Q	I	A	G	A	A	F	V	T	T	R	I	A	K	R	M	N	I	Q	N	L	W	S	R	P	Q	V	E	D	E	E	T	S	K	D	G	C	P	K	S	K	D	E	E	F	V	I	S	S	D	D	I	K	A	E					
NS4_Hu/GII.5/Saitama/T49/2001/JP	L	Q	I	A	G	A	A	F	V	T	S	R	M	T	R	R	L	N	V	S	S	I	I	W	V	R	P</																																		



ISSN 2052-2525

**Atomic structure and conformational variability of the  
HER2-trastuzumab-pertuzumab complex**

**Rémi Ruedas, Rémi Vuillemot, Thibault Tubiana, Jean-Marie Winter, Laura Pieri,  
Ana-Andreea Arteni, Camille Samson, Slavica Jonic, Magali Mathieu and Stéphane  
Bressanelli**

CONFIDENTIAL – NOT TO BE REPRODUCED, QUOTED NOR SHOWN TO OTHERS

SCIENTIFIC MANUSCRIPT

For review only.

**Category:** *research papers*

**Co-editor:**

*Dr F. Sun*

*Institute of Biophysics, Chinese Academy of Sciences, School of Life Science, University of Chinese Academy of  
Sciences, Datun Road 15, Chaoyang District, Beijing 100101, China*

*Email: feisun@bp.ac.cn*

**Submitting author:**

*Stéphane Bressanelli*

*CNRS, 1 avenue de la terrasse, Gif sur Yvette, Essonne, 91198, France*

*Email: stephane.bressanelli@i2bc.paris-saclay.fr*

---

# Atomic structure and conformational variability of the HER2-trastuzumab-pertuzumab complex

Rémi Ruedas\*#, Rémi Vuillemot°, Thibault Tubiana\*, Jean-Marie Winter<sup>+</sup>, Laura Pieri\*, Ana-Andreea Arteni\*, Camille Samson#, Slavica Jonic°, Magali Mathieu#, Stéphane Bressanelli\*

\* Université Paris-Saclay, CEA, CNRS - Institute for Integrative Biology of the Cell (I2BC), 91198, Gif-sur-Yvette, France.

# Sanofi, Integrated Drug Discovery, 13, quai Jules Guesde, 94403 Vitry-sur-Seine, France

<sup>+</sup> Nanolmaging Core Facility, Centre de Ressources et Recherches Technologiques (C2RT), Institut Pasteur, 75015 Paris, France.

° IMPMC-UMR 7590 CNRS, Sorbonne Université, Muséum National d'Histoire Naturelle, 75005 Paris, France

Single particle analysis (SPA) from cryogenic transmission electron microscopy (Cryo-EM) nowadays allows structural analysis in the native state of a broad range of proteins and macromolecular complexes. It is particularly attractive for complexes for which structure prediction remains intractable, such as antibody-antigen complexes. Here we describe a robust workflow for such complexes and the detailed structure of a particularly difficult complex involving two therapeutic antibodies. The workflow includes several validation points at pre-microscope, in-microscope, and post-microscope stages. The first part gives solid metrics about the biochemical properties of the samples through accessible biophysical measures. It is followed by a standard and efficient grid screening method to handle particle misdistribution and vitrification deterioration. The last part focuses on data processing. It highlights the strengths and limitations of current software as well as software currently under development in dealing with various kinds of heterogeneities, particularly continuous conformational heterogeneity. The workflow thus led to the structure of the complex between human epidermal growth factor receptor 2 (HER2) and the antigen-binding fragments from two distinct therapeutic antibodies binding to distant parts of the flexible HER2, pertuzumab and trastuzumab (HTP). Our HTP structure provides a more detailed view than the one previously available for this ternary complex, and, in addition, we describe the motions that can be extracted from our dataset. This allowed us to pinpoint a previously overlooked interaction that may contribute to explain the synergistic anticancer effect of the two antibodies. We further propose that the flexibility of the HTP complex, beyond the difficulties it causes for cryo-EM analysis, actually reflects regulation of HER2 signaling and its inhibition by therapeutic antibodies. Perhaps most importantly, we provide here a dataset for a smallish protein complex of 162 kDa for further development of software accounting for continuous conformational heterogeneity in cryo-EM images.

## Introduction

Human epidermal growth factor receptor 2 (HER2 or ErbB2) plays a critical role in cell signaling and its deregulation is involved in many cancers. It is a member of the ErbB family of tyrosine kinase receptors. There are four receptors in the family: epidermal growth factor (EGF) receptor 1 (also termed ErbB1/HER1), ErbB2/Neu/HER2, ErbB3/HER3 and ErbB4/HER4. ErbB receptors can form homo- and heterodimers

mediated by EGF-related peptide growth factors that bind to their extracellular domain (Olayioye et al. 2000). No ligand has been identified for HER2, and even if some papers suggest that it can form homodimers (Peckys et al. 2015)(Ghosh et al. 2011), it is more commonly described as the favored heterodimerization partner for the other ErbB receptors (Karunakaran et al. 1996) (Graus-Porta et al. 1997), (Ghosh et al. 2011). The heterodimers involving HER2 have higher signaling potency than heterodimers involving other ErbB receptors (Karunakaran et al. 1996). HER2-involving signaling is directly linked with cell growth and, when HER2 is overexpressed, with several cancers. Pro-oncogenic HER2 heterodimers with other HER receptors form through HER2 extracellular region (hereafter sHER2), leading to autophosphorylation of its intracellular region and downstream signaling. Dimerization and pro-oncogenic effects can be inhibited by antibodies directed against sHER2, prominently two monoclonal anti-sHER2 antibodies, trastuzumab (Herceptin) and pertuzumab (Perjeta). Trastuzumab binds to domain IV of sHER2 (see below) and was first approved by the Food and Drug Administration (FDA 1998) and by the European Medicine Agency (EMA) in 2000 for its use against HER2-positive breast cancer (EMA 2000). Fab (Fragment antigen binding) related effects of trastuzumab are manifold and include cytostatic effect, angiogenic factor regulation, antibody-mediated aggregation (Ben-Kasus et al. 2009) and internalization of HER2. Trastuzumab also inhibits a proteolytic cleavage of HER2 to a shorter form p59HER2 that is active in signaling (Molina et al. 2001). Its Fc related effects are antibody-dependent cellular cytotoxicity (ADCC) and IgG1 half-life increasing, the latter being important for favorable trastuzumab pharmacokinetics. However, acquired trastuzumab resistance has appeared over time, which needs to be overcome (Pohlmann et al. 2009). Pertuzumab has taken this role. Like trastuzumab, pertuzumab is a HER2 targeting humanized monoclonal antibody. It was first approved by the FDA in 2012 (FDA 2012) and by the EMA in 2013 (EMA 2013). Pertuzumab binds to domain II of sHER2 and its main Fab effect is to sterically block the HER2 dimerization process (Nami et al. 2018). Combination of trastuzumab and pertuzumab displays a synergistic effect (Nahta et al. 2004; Richard et al. 2016; Nami et al. 2018) and therefore is currently used as a very effective immunotherapy strategy against breast cancer (Scheuer et al. 2009; Blumenthal et al. 2013; Howie et al. 2019; Liu et al. 2022).

Accordingly, the molecular basis of trastuzumab and pertuzumab inhibition of HER2 has been the subject of much basic research. Particularly the atomic structures of complexes of sHER2 with Fabs have been reported (Table 1), first by X-ray crystallography for sHER2-trastuzumab (Cho et al. 2003) and for sHER2-pertuzumab (Franklin et al. 2004). In recent years HER2 structural biology has taken advantage of the ongoing revolution in cryogenic electron microscopy (cryo-EM) to provide atomic-level structures of large protein complexes. Thus, several important cryo-EM structures of HER2 complexes have been solved by single particle analysis (SPA), particularly the HER2-HER3 and HER2-HER3-trastuzumab (Fab) complexes (Diwanji et al. 2021). A major conclusion of these recent works is that dynamics are important for HER2 dimerization and for regulation of signaling (Diwanji et al. 2021; Bai et al. 2023). However, these

dynamics also make HER2 complexes difficult objects for SPA: Not only is sHER2 only 630 residues (70 kDa, not counting glycosylations), but its complexes tend to display several kinds of heterogeneity. This includes continuous conformational variability, that is currently the object of much research in the data processing cryo-EM community (Punjani and Fleet 2021; Vuillemot et al. 2023). Thus, the only structure of the tripartite complex between sHER2, trastuzumab Fab and pertuzumab Fab (hereafter called the HTP complex) is available only to lowish resolution (Table 1) and in particular does not allow visualizing to atomic resolution the HER2-trastuzumab interface, as this interface lies in the most flexible and least resolved part of HTP (Hao et al. 2019).

The architecture of HER extracellular regions includes four domains, consecutive in sequence, I to IV (Fig. 1a). In sHER2, domains I, II and III pack together while IV forms an elongated stem that connects to the transmembrane helix in full-length HER2. Accordingly, in maps of sHER2 but also of solubilized full-length HER2, the tip of IV is poorly defined and most residues beyond 570 are not resolved, including residues close to the trastuzumab paratope. Indeed, in cryo-EM SPA the most encountered limiting factors are particles distribution and orientation, ice thickness, and sample flexibility. (Passmore and Russo 2016; Vinothkumar and Henderson 2016; Rice et al. 2018; Weissenberger et al. 2021; Brown and Hanssen 2022). As to particles' distribution and orientation, there is a well-known and unpredictable gap between sample preparation and the vitrification process. Even if there is no foolproof method to predict the behavior of particles on the grids and during vitrification, an exhaustive quality assessment of the sample strongly increases success rates (Berrow et al. 2021; Raynal et al. 2021). Here we use such an approach to present a complete cryo-EM structure of the HTP complex that allows simultaneously resolving all HER2-trastuzumab and HER2-pertuzumab interactions. By using state-of-the-art image processing methods, we account for intrinsic flexibility of the complex and describe both its structure and motion.

## Results

### *An exhaustive biophysical view of the HTP complex*

The HTP complex was formed and purified by a protocol adapted from Hao et al, 2019, with a final size exclusion chromatography step to separate the full complex from free Fabs (see 'Materials and methods'). On the day freshly FPLC-purified samples were vitrified on grids, their integrity and suitability for cryo-EM were checked by UV-spectroscopy, dynamic light scattering (DLS) and differential scanning fluorimetry (nano-DSF) measurements. Good HTP samples were devoid of detectable chemical contamination (absorbance ratio 260nm/280nm were below 0.6) and had a low aggregation index ( $A_i = \frac{100 * A_{340}}{A_{280} - A_{340}} < 2$ ). The size variation induced by the binding of 2 fabs ( $\Delta 100$ kDa) on sHER2 was enough to induce a significant change in the diffusion coefficient measured by DLS and thus in the apparent hydrodynamic radius from 3.2nm ( $\pm 0.22$ ), 3.34nm ( $\pm 0.46$ ) and 4.61nm ( $\pm 0.11$ ) for trastuzumab Fab, pertuzumab Fab and

sHER2 respectively to 6.19 ( $\pm 0.34$ ) for the HTP ternary complex (Supp. Fig. 1). Systematic nano-DSF measurements gave direct information about the complex stability in solution through the melting temperature, that can be inferred from the variations in fluorescence ( $T_m$ ) or light scattering ( $T_{scatt}$ ) as temperature is increased. In isolation, both Fabs showed much higher light scattering counts and apparent size ( $1E^8$  counts,  $rh \sim 1000nm$ ) than sHER2 ( $6E^6$  counts;  $rh \sim 10nm$ ), indicating that Fabs aggregated upon denaturation (Supp. Fig. 1). While isolated sHER2 displayed one melting point ( $T_m$  67 °C), two melting points were observed for the HTP complex, confirming that the first (69 °C) and second (80 °C) melting points for HTP correspond to sHER2 and the Fabs, respectively. The  $T_m$  of sHER2 increased from 67°C in isolation to 69°C in the HTP complex, consistent with sHER2 stabilization by the two Fabs (Supp. Fig. 1).

The sample was further characterized by static light scattering coupled to size exclusion chromatography (SEC-SLS). The sample eluted at 0.3mg/ml (1.8  $\mu M$ ) had a mass of  $\sim 171$  kDa, indicating the HTP complex was formed and stable at this concentration (Supp. Fig. 1).

Mass-photometry was also performed on a frozen-thawed sample. Mass photometry is a single-molecule technique that applies the principle of interference reflection microscopy and interferometric scattering microscopy to quantify light scattered by molecules on a glass coverslip (Sonn-Segev et al. 2020). This experiment revealed five populations (Supp. Fig. 1.). The first three major peaks had apparent weights consistent with the HTP trimer, a "binary" complex probably formed by a single Fab complexed with sHER2 and a free population of Fabs. This result could be explained by the working concentration of 10 nM that could be close to the dissociation constants of the Fabs, since  $K_d$ s for the full antibodies were reported in the low nanomolar range by enzyme linked immunosorbent assay (1.43 nM for trastuzumab, 1.92 nM Pertuzumab) (Fu et al. 2014). These results are consistent with stability test performed by negative staining microscopy where frozen-thawed sample particles appear smaller than fresh sample particles (Supp. Fig. 2) but remain well distributed and concentrated. Finally, two higher weight peaks were also apparent, one barely visible at around 340 kDa and the other a substantial part of the counts (23%) at 685 kDa. The latter one is also visible by DLS, as a sharp increase in polydispersity in frozen-thawed samples from a low value of the polydispersity index of  $0,07(\pm 0,02)$  for the fresh sample (Supp. Fig. 2). All these results from DLS, mass photometry and negative staining indicate that freeze-thaw cycles tend to induce dissociation and/or aggregation of HTP. Thus cryo-EM grids were prepared only from fresh samples and at concentrations much higher than the dissociation constants of the Fabs.

### **Grid type strongly impacts particles behavior and proved to be the critical factor to reach high resolution in cryo-EM SPA**

HTP samples at a typical concentration of 0.1-0.3 mg/ml were applied on grids and

plunge-frozen (see Materials and Methods and Table 3). Grids screening was performed according to hierarchical criteria. First the grids overall quality (*e.g.* grid and ice integrity and a good thickness gradient along the grid), then the ice quality including amorphous state and relative thickness, finally the particles' distribution in thin ice. The HTP particle is too small for integrity and/or orientation bias to be checked by direct visualization. Thus, the only visual quality criteria were the particles' density and distribution over the holes. Our grid screening strategy is schematized in supplementary figure 3.

Data were collected and processed as described in Material and methods and table 3. The major differences according to grid type are summarized in supplementary figure 4. According to sample characterization, there was no reason to think that aggregation would be a concern with a freshly prepared sample. However, when we used the published conditions for cryo-EM grid preparation of HTP on quantifoil grids (Hao et al. 2019), we experienced difficulties in reproducibly obtaining well-dispersed complexes in thin ice. Indeed, when using quantifoil grids, particle clustering repeatedly and vexingly appeared upon blotting and plunge-freezing grids despite numerous biochemical screening trials (Table 2 & Supp. Fig. 4). A low-resolution reconstruction of the HTP trimer could still be made from such a grid. However, 3D classification revealed that besides the full HTP trimer, there was a subpopulation where trastuzumab was missing. Further exhaustive image analysis using deep particle picking revealed that the HTP complex is furthermore also present as a doublet occurring by self-association through two parallel pertuzumab Fabs (Supp. Fig. 4). Map quality for the doublet does not allow to precisely identify which amino acids are involved in this self-association, but it is good enough to fit atomic models that show that i) pertuzumab-pertuzumab contacts occur through the variable regions, but opposite the pertuzumab paratope and ii) the elbow angle (angle between variable and constant regions) is different from the (average) one in the free HTP complex. This doublet could be one of the building blocks of the clusters that are the main factor impairing high resolution reconstruction in quantifoil grids.

We therefore tested gold grids called Ultraufoil. No aggregation/clustering are visible on those grids, but partially assembled complex forms are present (Supp. Fig. 4). Four different objects can be reconstructed. Two of them can be unequivocally attributed to the full ternary complex HTP and the partially assembled complex pertuzumab-sHER2, as on the classic quantifoil grid. The third and fourth volumes appear smaller and could correspond to a partially assembled complex trastuzumab-sHER2 and unbound Fabs or noise. Thus, Ultraufoil grids helped with the HTP complex' dispersion in ice but not with its integrity. At any rate, this in-depth analysis shows that objects detected after freeze-thawing by mass photometry (both incomplete complexes and doublets) and DLS (higher order aggregates) can actually be found in ice after plunge-freezing a fresh sample (Supp. Fig. 2 & 4).

In view of this we next played with different factors that could further improve sample



quality upon vitrification. This included grid material but also grid treatment, varying biochemical conditions and blotting time and force prior to plunge-freezing (Table 2). Each condition was tested in duplicate for a total of some eighty grids. Using ultrathin carbon coated (UTC) quantifoil grids was the critical factor that finally allowed to have undamaged and well-distributed HTP particles in thin ice (Supp. Fig. 4, lower panel). We could thus obtain a 3D reconstruction to an overall resolution of about 3.2 Å, compared to the previous 4.4 Å (Hao et al. 2019).

### **Complete, simultaneous atomic description of both HTP epitope-paratope interfaces: a fourth loop in the trastuzumab paratope**

The map clearly shows that the standard image analysis protocol used (hereafter the baseline protocol, see methods) allows accurate reconstruction of the I to III part of sHER2 and the variable region of the bound Pertuzumab Fab (both the part coming from the heavy chain  $V_H$  and from the light chain  $V_L$ ) (Fig. 1a). However, the map is of much lower quality in the region comprising sHER2 IV and the trastuzumab Fab. We further ran two focused refinements, one in a mask including sHER2 domains I to III and the pertuzumab Fab (Fig. 1b), and the other including most of sHER2 domain IV and the trastuzumab Fab (Fig. 1c) and combined the two focused maps (Fig. 1d). Focused refinement further improved the map even in the pertuzumab region, with a local resolution reaching 2.5 Å almost throughout this whole subpart of the combined map (Fig. 1b). In fact, the interface between HER2 and pertuzumab is better defined than in the available crystal structure (Table 1). We can thus ascertain that this interface is the same as in that former structure (Fig. 2b) (Franklin et al. 2004), with only minor changes. This confirms that simultaneous binding of trastuzumab does not affect the HER2-pertuzumab interface, as previously reported (Hao et al. 2019). Focused refinement particularly improved the trastuzumab region (Fig. 1c). The map is very clear in the HER2-trastuzumab interface. To describe it, we use the same numbering as in the initial sHER2-trastuzumab crystal structure (Cho et al. 2003), *i.e.* starting the sHER2 sequence at residue 1 and not counting the 22-residue signal peptide of the HER2 ORF. The HER2-trastuzumab interface thus involves at its center the hydrophobic HER2 bulge 570-573 and on either side loop 557-561 and some residues of 593-603, in agreement with the crystal structure (Fig. 2a, yellow). However in this region our maps also show clear features of a fourth, long HER2 loop 581-590 involved in the interface (Fig. 1, Fig. 2a, purple). Particularly residues D585-S587 at its tip make contact to a trastuzumab surface contributed by several CDR's (complementarity-determining regions as defined in Ling et al. (Ling et al. 2018)): a cleft of the heavy chain at the bottom of which is R98 and lined by CDR's H3 and H1 and an aromatic patch provided by the light chain CDR L2 (Fig. 2a). Density in this region was actually present if barely visible in Hao et al. and Diwanji et al. cryoEM maps, supporting our decision to redefine the HER2 trastuzumab binding surface as a four-part epitope (Fig. 2).

### **Continuous conformational variability analysis**

The combined focused map is still less good in the Trastuzumab region than in the Pertuzumab region (Fig. 1d). This may be due in part to the fact that Pertuzumab Fab together with domains I, II and III of HER2 account for 941 amino acids (61.5% of the complex) against only 587 amino acids (38,5%) for HER2 domain IV and Trastuzumab Fab, making the smaller region more difficult to align. But this is also the most flexible region of the HTP complex and is likely not well described by a single conformation. Indeed, 3D classification without alignment (that is, using the alignments computed in the baseline refinement, see Material and methods) converged to roughly equally populated classes with distinct conformations and no incomplete class (not shown, we tried up to ten classes), suggesting that heterogeneity in the "UTC quantifoil" dataset is continuous rather than discrete. We therefore applied specific image processing protocols that seek to account for continuous conformational variability in cryo-EM images.

### 3DFlex from cryoSPARC

We thus used the newly released 3D Flexible Refinement (3DFlex)(Punjani and Fleet 2021) of cryoSPARC (Punjani et al. 2017). 3DFlex, a motion-based deep neural network model of continuous heterogeneity, extracts a latent space of continuous motion from 2D image data to learn non-rigid motions. The outputs of 3DFlex are on the one hand series of volumes describing the motions and on the other hand a single 'canonical' reconstructed map from 'back-deformed' images to recover high-resolution details of flexible parts. As the current 3DFlex beta version is demanding on computational resources, we first applied 3DFlex on a 286,008 particles subset (hereafter 286k subset) of the full 711,008 particles dataset previously obtained as described in Material and methods. As we had a few strongly favored orientations in the dataset, we did not select a random subset but made a python script RSPFilter, to specifically remove particles in the most favored orientations while keeping all particles from less populated orientations (Supp. Fig. 5). The 3DFlex canonical map reconstructed from this subset of particles shows a marked improvement in the trastuzumab region compared to the map obtained with the baseline protocol (Figure 3a). This suggests that the latent space extracted by 3DFlex is correct and thus that it was successful in capturing the conformational landscape in the 286k subset. Indeed, the main motions generated by the 3DFlex flow generator show several major modes that are consistent with the expected dynamics of HER2 and Fabs (Supp. Data movie 1). Thus, the three main motions are located, with increasing amplitude and complexity, in the hinge between the constant and variable regions of pertuzumab, in the equivalent hinge of trastuzumab, and in HER2 domain IV (Figure 3c and supp. movies). It is particularly noteworthy that despite considerable motion in IV, the four-part trastuzumab epitope remains associated, seemingly without change, with the long IV loop 581-590 engaged to Trastuzumab in all conformations. Thus the apparent disorder that had precluded its identification as part of the epitope in the initial crystal structure (Cho *et al.*, 2003) does not seem to be due to local flexibility of this loop. To better ascertain these points, we next applied the same 3DFlex protocol to the full 711,008 particle dataset. However,

the additional particles led to a dampening of the motions generated by 3DFlex, though the motions were qualitatively the same as output for the 286k RSPFiltered subset. We tried changing several different parameters such as 2 to 5 latent dimensions during Flex train, initializing with latent coordinates from 3D Variability Analysis, using a segmentation file during Flex mesh prep, or cropping and downsampling to different extents during Flex data prep. The only parameter that had some effect was the extent of cropping applied prior to Flex train, as tighter boxes generally led to less dampened motions. For instance, cropping the box from 230pix (266.3 Å) to 200 pix during Flex data prep increased the amplitude of movements both for the 286k subset and the full 711k dataset. Nevertheless, the resulting movements were still dampened for the full dataset. The largest distance between extremes in conformations was in the constant region of trastuzumab and reached ~21 Å for the RSPFiltered 286k dataset (Figure 3c) while it remained below 5 Å for the full dataset. Accordingly, although the 3DFlex canonical map presents much better features in the less mobile pertuzumab region with the full dataset than with the 286k subset, it is actually degraded in the trastuzumab region (Figure 3a-b).

## MDSPACE

3DFlex has the advantage that it learns the continuous conformational variability directly from the dataset, but the disadvantage that prior knowledge in the form of an atomic model cannot be input or guide the search, even in such a difficult case as HTP. To probe the validity of the 3DFlex results and better assess the continuous conformational variability of the HTP complex, we therefore used a different approach, the recently published MDSPACE (Vuillemot et al. 2023). Molecular Dynamics simulation for Single Particle Analysis of Continuous Conformational heterogeneity (MDSPACE) uses an atomic model of the complex as input (here, the model built from the data) and applies a per-particle flexible fitting using MD simulations to match the conformation in each particle image. The flexible fitting uses and normal mode empowered MD simulation to accelerate the process in an iterative scheme that helps to extract the conformations from projections where the flexibility is more difficult to discern (particularly particles viewed along the direction of one of the main movements). A coarse-grained or atomic-level conformation is thus assigned to each particle in the dataset. As shown in Figure 3D, using the whole dataset and a coarse-grained HTP model, the MDSPACE analysis identified major movements that are very similar to those resulting from the 3DFlex analysis, particularly in the HER2/trastuzumab region. However, the movements from MDSPACE are of greater amplitude (47 Å between extremes) than from 3DFlex, even when comparing with the 286k subset 3DFlex analysis (compare Fig. 3D with Fig. 3C).

Despite the less restricted conformational space described, the MDSPACE coarse-grained models do not show variation of the HER2-trastuzumab interface. To further probe this point, we ran a second MDSPACE analysis using an all-atom HTP model on a subset of 5k particles. Aligning the resulting all-atom models on trastuzumab shows

that the main local motion in this region is around the Fab's elbow angle between the constant and variable regions (Fig. 3E, left) and that the interface with HER2 remains stable (Fig. 3E, right). Thus, continuous conformational variability analysis consistently shows the stability of the fourth part of the HER2-trastuzumab interface (loop 581-590 with CDR's H3, L1 and L2).

## Discussion

Our study sheds further light on the simultaneous molecular interactions of trastuzumab and pertuzumab with HER2. We propose the involvement of a fourth interaction point between trastuzumab and HER2 domain IV. This fourth part of the sHER2 trastuzumab epitope, loop 581-590, had actually been built in the lower resolution map of the HTP complex, albeit in a different conformation (PDB 6oge), but it is not mentioned in the paper (Hao et al. 2019). Our flexibility analysis shows that loop 581-590 is an integral part of the trastuzumab-HER2 interface in all distinguishable conformations of the HTP complex. This loop completes the striking likeness between the epitopes of trastuzumab in domain IV and pertuzumab in domain II. In the HER2 modular architecture, domain IV is a repeat of domain II and loop 581-590 in domain IV is the counterpart of the "dimerization arm" in domain II. Both contain hydrophobic bulges that in the case of the dimerization arm is known to be involved in the heterodimerization process (Garrett et al. 2002; Franklin et al. 2004; Diwanji et al. 2021; Bai et al. 2023). The close resemblance between the binding sites of trastuzumab and pertuzumab suggests a common mechanism of binding, while their synergistic effect on HER2 signaling (Nahta et al. 2004; Richard et al. 2016; Nami et al. 2018) hints that different processes are inhibited by the two antibodies. Interestingly, although there is no evidence of direct contact, the domain IV loop in HER2 is near the domain IV loop in HER3 in the HER2-HER3 complex and it is displaced upon binding of trastuzumab in the HER2-HER3-trastuzumab complex, where HER2 and HER3 remain associated through the dimerization arm (Diwanji et al. 2021). This may hint at a signaling mechanism in which both the 581-590 loop in domain IV and the dimerization arm in domain II are required.

Our study supports the typical workflow and troubleshooting in cryo-EM, especially the gap between the "on the bench" sample preparation and characterization and finding a suitable vitrification process avoiding the twin problems of dissociation and aggregation. It can provide a workflow for new cryo-EM users. Although the current cryo-EM software allows a rapid data processing, we would like to emphasize particularly the advantages of a solid sample biophysical characterization prior to or concurrently with the cryo-EM steps. Not only does it give a useful (though never foolproof) indication of the most likely sample conditions for good vitrification, but it also allows a better understanding of the observations in the cryogenic electron microscope. In the case of the HTP complex, a particularly difficult complex for cryo-EM, we show that fine characterization by mass photometry, DLS and negative staining EM observations of frozen/thawed or aged samples actually detect the dissociated or

aggregated states that occur in unsatisfactory vitrification processes. We conclude that they could mimic an accelerated aging of the sample. We also confirm that the stability and dispersion of the biological sample during vitrification is intricately linked to the grid material and find, by trial and error, that an ultrathin carbon coating on the grid is the crucial factor for a satisfactory vitrification of the HTP complex. Many factors can impact the behavior of a sample on the grid, just as is the case for crystallization conditions in crystallography. Seemingly minor changes such as the brand of any reagent can be critical. Therefore, it is important to minimize processing, handling, and times between the output of the last purification step (usually size exclusion chromatography) and vitrification. In the case of the HTP complex, despite a high variability from one sample to another, an integrative approach helped to bridge the gap between the bench and the vitrification.

Finally, we make available to the community a dataset that offers several advantages for further method development in protein studies by cryo-EM. It is of a smallish, compositionally homogeneous object with a distinct shape, large continuous conformational variability, and a strong orientation bias while still sampling all orientations. We provide a user-friendly utility, RSPFilter, to analyze and mitigate the effects of orientation bias in cryo-EM data and a concrete example of its usefulness and interest. Using this dataset, we show that two state-of-the-art software packages for conformational variability analysis perform differently in the challenges posed by this dataset. Particularly MDSPACE performs a finer analysis as it detects a wider range of motions, likely because it makes use of additional structural information (the atomic model built from the data). The smaller box size of our dataset compared to those typically used to benchmark such software offers a double advantage of less computational demand and the possibility of obtaining the full resolution for both images and atomic models. This makes it an ideal test case for further development of cryo-EM data processing software accounting for flexibility. Overall, our findings highlight the potential of our dataset for advancing methodology and software development in cryo-EM studies of flexible protein complexes.

## Materials and Methods

### Sample preparation

**Human Her2 / ErbB2 ectodomain** comprising residues Thr23-Thr652 of the HER2 open reading frame (numbered 1-627 in this study) followed by his tag was purchased as a lyophilized powder from ACROBIOSYSTEMS Inc (Cat. No. #HE2-H5212). It was resuspended in 10mM HEPES pH 7.5, 150mM NaCl and injected on a Superdex 200 increase 10/300 GL size exclusion column at 7°C. Fractions containing sHER2 were concentrated up to DO=3.2, aliquoted in 15 µL fractions, fast frozen in liquid nitrogen and stored at -80°C.

**Monoclonal antibodies (mABs) pertuzumab and trastuzumab** purchased from CliniSciences (Nanterre, France, Cat. No. HY-P9912 and HY-P9907) according to the

supplier Pertuzumab and trastuzumab are recombinant humanized monoclonal antibody produced by recombinant DNA technology in a mammalian cell (Chinese Hamster Ovary) culture containing the antibiotic, gentamicin. Pertuzumab and Trastuzumab have an approximate molecular weight of 148 kDa. They were solubilized in sodium phosphate buffer at pH 7.5 at ~3mg/ml (DO~4.2) then aliquoted and stored at -20°C before digestion using commercial immobilized papain (#20341 Thermo Fisher Scientific) to produce Fabs. For digestion, 250 µL of 50% immobilized papain were poured in a 1ml spin column (#69725 Thermo Fisher Scientific) and washed with ultrapure water using a benchtop centrifuge. Then the papain was activated and equilibrated with the digestion buffer containing 20mM cysteine, 10 mM EDTA, 20 mM sodium phosphate pH 7.5. Monoclonal antibodies were complemented up to 20 mM cysteine. Then the mAb was digested at 37°C during 4-5 hours under constant agitation using Intelli-mixer to avoid papain beads sedimentation. After digestion, elution was done by several centrifugations at 1000g on an Eppendorf centrifuge. Fabs were purified from Fc and undigested IgG by ion exchange chromatography using 3 ml SP-Sepharose resin (#17-0729-10 GE-Healthcare). Before injection, Fabs were dialysed overnight at 7°C in 50mM Acetate buffer pH4 then injected in the SP-sepharose. Loading and washing steps were done in buffer A (50mM Acetate buffer pH4), elution was performed with a gradient of 40% buffer B (Buffer A + 1M NaCl) in 120 min at 0.3 ml/min. The first peak contains Fab, the second peak contains Fc and undigested IgG. Fractions containing the Fabs were neutralized with 1 M HEPES pH 7.5. Then they were buffer exchanged with 10 mM HEPES pH 7.5 and 150 mM NaCl by overnight dialysis at 7°C or by fast buffer exchange using centricon 30 kDa. They were concentrated up to ~1 ml and injected on a Superdex 200 10/300 GL size exclusion chromatography (SEC Cytiva). Two SDS-PAGE 14% acrylamide with fractions of all steps were performed with and without reducing agent to visualize papain digestion and the purity of samples. Then they were concentrated up to DO=3.2, divided in 15 µL aliquots, fast frozen in liquid nitrogen and stored at -80°C.

**Complex formation.** The complex was formed by incubating the two Fab's (from trastuzumab and pertuzumab) with sHER2 on ice during a few minutes. The ratio was 1:1.2:1.2 sHER2/Fab1/Fab2. The complex was then purified through a S200 GE healthcare size exclusion chromatography in 10 mM HEPES pH 7.5 and 150 mM NaCl. A shift of peak corresponding to an increasing size of the object was obtained compared to the sHER2 alone and Fab alone injections.

### Biophysical measurements

**UV spectrum.** 2.5 µL of the sample was deposited on the cell of a nano-drop spectrophotometer (nanodrop one Thermo Fisher Scientific) and a UV spectrum was recorded from 230nm to 350nm. Quality assessment is done as described in Raynal et al. 2021 (Raynal et al. 2021)

**DLS & nano-DSF.** Measures were performed on the Prometheus Panta from

nanotemper. Samples were diluted to an OD 280 nm signal between 0.3 and 1 and centrifugated a 11 000g during 10 min. 10  $\mu$ L were used to fill the capillary of the system and start the measurement. DLS measurement was performed at 15°C. DSF was measured on a ramp from 15°C to 95°C with steps of 0.5°C. Data were processed using the nano-temper software.

**Mass photometry.** Mass and relative abundance of samples were measured using a TwoMP mass photometer (Refeyn Inc., Oxford, UK). A frozen 1 month old ternary complex was thawed prior to deposition in the Refeyn cell. Sample was loaded on a silicone gasket (CultureWell™ reusable gasket, 3 mm diameter  $\times$  1 mm depth, Grace Bio-Labs) and placed onto clean glass coverslips (Refeyn Inc.). In each well, 10  $\mu$ L of buffer (10mM MES pH6.5, 150 mM NaCl) was deposited and a blank measurement made. Subsequently, protein sample was mixed with the buffer in the well to a final concentration of  $\sim$ 10 nM. The molecular weight of the samples was assigned by contrast comparison of known mass standard calibrants measured on the same day.

**SEC-SLS.** Complex was prepared as described above and injected on a OMNISEC (Malvern Panalytical) with a Superdex 200 10/300 GL equilibrated in 10 mM HEPES pH 7.5 and 150 mM NaCl coupled with a static light scattering (SLS) detector.

**Negative Staining.** Freshly made complex was diluted 100 times and a 3 $\mu$ L drop placed on the grid. After 15 seconds the drop was dried with Whatman paper and then colored using 2% uranyl acetate. Images were acquired with a Tecnai Spirit 120 kV LaB6 (FEI) equipped with a K2 Base Direct Detection Camera (Gatan).

### Cryo-EM grid preparation

Immediately before grid preparation, the grids were glow-discharged using GloQube plus system. Except for the ionization tests performed on quantifoils (table 2), the routine protocol for Quantifoil, UltraAufoil, UTC Quantifoil and C-flat grids is 10-20 sec at 12 mA under vacuum. All grids were prepared by plunge freezing in liquid ethane using Vitrobot Mark IV system (Thermo Fisher Scientific, USA). Blotting parameters were adjusted for each session and depend on the vitrobot used. Usually, blotting time was near 4 seconds, blotting force 0 with no waiting time before blotting. Sample deposited was 2-3  $\mu$ L with an OD 280 nm between 0.1 and 0.3 (table 3), in a humidity chamber kept at 100% for all grid preparation. The temperature did not impact the sample distribution on the grid, consistent with DSF and DLS measurements, and so the grids described in this paper were prepared at 15°C. During the procedure, sample was kept on ice. For the exhaustive data processing only the grids Quantifoil 1.2/1.3 $\mu$ m, UltraAufoil 0.6/1 $\mu$ m and UTC+quantifoil 1.2/1.3 $\mu$ m were used.

### Cryo-EM data acquisition

All data collection parameters are summarized in table 3. The classic-Quantifoil and UTC-Quantifoil grid movies were recorded in eer format, UltraAufoil in tiff.

## Data Processing

**UTC Quantifoil.** Processing is summarized in supplementary figure 6. Data were processed using Cryosparc v. 3.3.2 and v. 4.1.2 with 3DFlex. Eer movies format were imported with 54 eer fractions and an upsampling factor of 1.

**Baseline protocol:** All movies were motion corrected using patch motion correction multi with a F-crop factor of 0.5, using all frames (total dose 60  $\text{e}^-/\text{\AA}^2$ ). Then CTF estimation was done using Patch CTF estimation with default parameters. Micrographs were screened using curate exposure according to CTFfit, total full frame motion distance and average defocus values. Bad ice (cubic or hexagonal phase), blurred or empty images were then discarded. 2D templates for picking were obtained by a first automatic (blob) picking over 50 random micrographs followed by 2D classifications. A few good classes were used for template picking with a particle diameter of 160  $\text{\AA}$  and a minimum distance of 0.5. Coordinates were screened using inspect particle picks and the selected 2 954 825 particles were extracted in a box size on 280 pixels (i.e. 313.6  $\text{\AA}$ ). After 2D classification, ab-initio volume generation was performed separately with particles from good and from bad 2D classes. The two 'good' volumes and three 'bad' volumes were used for successive rounds of heterogeneous refinement with increasing resolution, interspersed with homogeneous refinement of the best class. Particles were re-extracted once to recenter them in the box based on an intermediate homogeneous refinement (supp. Fig. 6). This procedure yielded 742 378 particles in the best class. These particles were 2D-classified again and particles from noisy classes were discarded. The remaining 711 008 particles were CTF global and local refined in cryosparc, before being injected to a new homogeneous refinement yielding a map at an overall 3.3  $\text{\AA}$  resolution.

**Local refinement:** From this map two masks were created using ChimeraX and CryoSPARC map tools, one mask covering the pertuzumab Fab - HER2 domains I-II-III region and the other covering the trastuzumab Fab - HER2 domain IV region. Two distinct local refinements were done using those masks and all particles without re-splitting the particles into two gold standard halves.

**Composite map building:** Using Phenix software, the two local refinement maps were merged to create a composite map. We used the global homogeneous refinement map from the baseline protocol as input reference, a preliminary pdb model of the complex with two atom selections corresponding to the two masks of the local refinements (pertuzumab/sHER2 and trastuzumab/sHER2 regions, respectively), and then the map and half-maps for each local refinement. The resolution was set at 2.5  $\text{\AA}$ . The output of the algorithm is a composite map and the corresponding half-maps. These maps were converted into mrc format by chimeraX and then imported into cryosparc to calculate the GS-FSC and local resolution.

**RSPFilter.** We made a Python 3.8 script RSPFilter to analyze and filter a set of particles described in a STAR file. Particles are filtered using a grid density technique to create a



more homogeneous repartition of the particles according to two properties selected in the STAR file. The star file Python package reads a STAR file and converts it into a Pandas Dataframe. The density (i.e. number of particles per square) is computed with the "binned\_statistic\_2D" function of Scipy package 1.9.3 (Virtanen et al. 2020) In each square, if the density is above a chosen density cutoff, particles are randomly eliminated until the density corresponds to the cutoff. Graphics are generated with the plotly package (Plotly 2015) and the web interface with Panel 0.14. RSPFilter is available on github (<https://github.com/tubiana/RSPFilter>), (Tubiana 2022) and can be installed through pypi.

In the case of UTC quantifoil data, we first converted the Cryosparc files "\*\_particles.cs" and "\*\_passthrough\_particles.cs" from cs format to *star* format with "csparc2star.py" from the pyem collection (Asarnow et al. 2019). Then using RSPFilter we computed the two-dimensional particle density according to the rotation angle ("rlnAngleRot") and the tilt angle ("rlnAngleTilt") with a resolution of 3°. A cutoff of 254 yielded the 286k particle subset.

**3DFlex.** We ran the Cryosparc v. 4.1.2 3DFlex analysis on the full set of particles and on the 286k subset. The protocol for the results presented here involved cropping from 280 to 220 pixels and downsampling to 128 pixels in Flex data prep, adjusting mask threshold and dilation in Flex mesh prep to have a mesh generously covering the HTP complex map considering its potential flexibility (Threshold 0.1 and a mask dilation set at 5 Å for the 286k subset) and 5 latent dimensions in Flex train. Flex reconstruct was run with 60 BFGS iterations and force re-do GS split.

**MDSPACE.** The analysis of MDSPACE (Vuillemot et al. 2023) was performed on particles that were cropped from 280 to 220 pixels and binned by 2 (final pixel size of 2.08 Å<sup>2</sup>). Two different forcefields were used for the MD simulations in MDSPACE. The first is a coarse-grained Gō model including C $\alpha$  atoms only, that was used to analyze the entire dataset of full set of particles and study the global variability. The second is the all-atom CHARMM forcefield which was used to analyze a random subset of 5k particles and study the local variability. In both cases, we used the atomic model derived from the combined map obtained with the two focused refinements as the initial conformation for the fitting. We also used the particle pose obtained by Cryosparc as the initial particle pose used for MDSPACE. The model was energy minimized with each forcefield and the simulations were performed for 200 picoseconds with a time step of 2 femtoseconds using GENESIS (Kobayashi et al. 2017). The MD simulations were empowered using Normal Mode Analysis (NMA) through the NMMD (Vuillemot et al. 2022) to accelerate the fitting. The normal modes were computed from the initial conformation using an elastic network model (ENM) as described in (Vuillemot et al. 2023) and the four lowest frequency normal modes (mode 7 to 10 as the first six modes corresponds to rigid motions) were selected for the fitting. In the analysis of the full set of particles dataset using C $\alpha$ -Gō model, MDSPACE was run for 2 iterations, with a first iteration (training) including a random subset of 1k particles, and the second iteration

including the entire dataset and using the three principal motions learned from the previous iteration calculated by principal component analysis (PCA), through the iterative protocol described in the Vuillemot et al. 2013 paper. The fitted structures were rigid-body aligned with respect to the pertuzumab Fab of the initial conformation and projected onto a low-dimensional space using PCA. The low-dimensional space was used to obtain the presented averaged structures and the 3D reconstructions. For that, semi-automated clustering was performed on the PCA space by using the MDSPACE animation tool in Scipion (Rosa-Trevín et al. 2016), along the first three PCA components. The presented averaged structures were obtained by averaging the atomic coordinates of the fitted structures that belong to each cluster. The presented 3D reconstructions were obtained by performing 3D reconstruction on the particles contained in each cluster. In the analysis of 5k particles using CHARMM forcefield, MDSPACE was run for one single iteration. The fitted structures were rigid body aligned to the trastuzumab Fab and the atoms of the trastuzumab Fab and its interface were selected to calculate the root mean square fluctuation (RMSF) that measures the average variation of each residue of the 5000 fitted structures. The entire MDSPACE workflow was executed in the ContinuousFlex plugin (Harastani et al. 2022) of Scipion (Rosa-Trevín et al. 2016).

**Classic Quantifoil data processing.** Two sets of eer movies (5566 and 5958) collected on the same microscope using the same data collection parameters were merged. Data were processed first with the baseline protocol with minor variations. The pixel size was 0,6 Å and 8 870 micrographs were kept after screening. From 683 895 particles found, 2D classification revealed the presence of a doublet of HTP complex in addition to HTP-sized particles. Thus, we used the deep picker topaz (Bepler et al. 2019) to selectively find the doublet particles over the dataset. From 2D classification, a template showing the doublet was used for template picking over 1000 micrographs. The 79 594 extracted particles were Fourier downsampled by 2 (pixel size 1,12 Å) and screened by two successive 2D classifications retaining only classes showing a doublet. The remaining 1 286 particles and their associated micrographs were selected and used for topaz train. Training was done with a downsampling factor of 6 (effective pixel size of 3,6 Å) and an expected number of 10 particles per micrographs. Other parameters were kept as default. Topaz extract was performed over the whole set of screened micrographs with a radius of extracted regions set to 35 pixels. 151 242 particles were found and classified into three 3D classes using heterogeneous refinement and a pixel size of 2,5 Å (Voxel 144). Particles assigned to the doublet class were combined with those obtained with template picking. Duplicates were removed using a minimum separation distance of 50 Å.

A last round of heterogeneous refinement of the 755 345 particles with 4 volumes (HTP, Doublet, pertuzumab/sHER2 and trash/junk volume) with a pixel size of 4,7 Å (Voxel 72) followed by a NU-refinements for each class yield the maps shown in supplementary figure 4.

**UltraAfoil data processing.** A set of 5 637 tiff movies format with a pixel size of 0.86 Å were imported and process as describe in the baseline protocol except for minor variations. Template picking over all the remaining 4098 screened micrographs, led to 1 014 383 particles after an 'inspect particle picks' correction. from a that were extracted within a box size of 360 pixels binned to 120 pixels (effective pixel size of 2.58 Å). After 2D classification, ab-initio volume generation was performed separately with particles from good and from bad 2D classes. Four 'good' volumes and two 'bad' volumes were used for three successive rounds of heterogeneous refinement with increasing resolution removing the particles from the 'junk' classes. 740 544 particles were kept for a last round of heterogeneous refinement that give the 4 volumes shown in supplementary figure 4.

**Model building.** To initiate the reconstruction, we combined the pdb files from 1N8Z (trastuzumab) and 1S78 (pertuzumab) removing HER2 from 1S78. We built a *de-novo* extra cellular domain of HER2 using alphafold according to the sequence used (Thr23-Thr652). In Pymol, we aligned locally the alphafold model on the 1N8Z and completed the missing part with the alphafold model. The regions that were completed are Leu101-Ser111, Ala302-Thr306, Asp360-Ser365, and Ser580-Ile591, Asn607-Thr652.

**Model refinement.** The initial model was first fit in the density of the composite map and briefly overall refined using Isolde (Croll 2018) in ChimeraX (Pettersen et al. 2021). Then the model was refined in coot and phenix using alternatively the composite map from the two focused refinements and the global homogeneous refinement map at the end of the baseline protocol. The final phenix real space refinements were done in the global homogeneous refinement map.

## DATA AVAILABILITY

The cryo-EM atomic model, maps used in refinement (global map, focused maps and composite map), and images have been deposited in the Protein Data Bank (ID code 8Q6J, 8PWH), Electron Microscopy Data Base, <https://www.ebi.ac.uk/emdb/> (EMD-18188, EMD-18189, EMD-18190, EMD- EMD-17993), and Electron Microscopy Public Image Archive, <https://www.ebi.ac.uk/empair/>.

## ACKNOWLEDGEMENTS

We wish to thank Malika Ould Ali from the I2BC Cryo-EM facility as well as Maelenn Chevreuil and Bertrand Raynal, from the Institut Pasteur, plateforme de biophysique moléculaire of C2RT. We acknowledge the cryo-EM platform nanoimaging of the

institut Pasteur.

## FUNDING INFORMATION

This work benefited from the CryoEM platform of I2BC, supported by the French Infrastructure for Integrated Structural Biology (FRISBI) [ANR-10-INSB-05-05] and from support of MOSBRI (European Union's Horizon 2020 research and innovation program under grant agreement No 101004806).

TT is the recipient of an ANRS-MIE postdoctoral fellowship ECTZ189696.

We acknowledge the support of the French National Research Agency - ANR (ANR-19-CE11-0008 to SJ) and access to HPC resources of IDRIS granted by GENCI (AD010710998R1 and AD010710998R2 to SJ).

## Figures and tables

### Tables

PDB code	Description	Technique	Res. (Å)	Date	Publication & authors
1s78	HER2/Pertuz	X-ray	3.25	2004	Insights into ErbB signaling from the structure of the ErbB2-pertuzumab complex (Franklin et al. 2004)
1n8z	HER2/Tastuz	X-ray	2.52	2003	Structure of the Extracellular Region of HER2 Alone and in Complex with the Herceptin Fab (Cho et al. 2003)
7mn8	HER2/HER3/ NRG1 $\beta$ Trastuz	CryoEM	3.45	2021	Structures of the HER2-HER3-NRG1 $\beta$ complex reveal a dynamic dimer interface. (Diwanji et al. 2021)
6oge	HER2/Trastuz /Pertuz	CryoEM	4.36	2019	Cryo-EM structure of Her2 extracellular domain-Trastuzumab Fab-Pertuzumab Fab complex (Hao et al. 2019)
6bgt	HER2/Trastuz mutant	X-ray	2.7	2017	To be published ( <i>Christie, M., Christ, D. not published</i> )
3be1	HER2/Trastuz (dual)	X-ray	2.9	2009	Dual specific bH1 Fab in complex with the extracellular domain of HER2/ErbB-2 (Bostrom, J.M., Wiesmann, C., Appleton, B.A.)

**Table 1.** State of the art. All structures of HER2/Fab complexes available in the pdb.

<b>Overview of tested parameters</b>			
<b>Parameters</b>		<b>Part. dist.</b>	
<b>Grid treatment</b>	5mM MgAc post-plasma, 10% Acetone, 10% EtOH, 10% SDS.		<b>Grids physics</b>
<b>Blotting T°</b>	4°C, 15°C, 22°C		
<b>Blotting parameters</b>	Blotting times, blotting force, waiting time		
<b>Plasma cleaner</b>	5-50mA 0-40sec		
<b>Concentration</b>	DO~0.3	Partial success	<b>Biochemical condition</b>
<b>pH</b>	4,3 – 7,5 - 9	Fail	
<b>Salt</b>	NaCl 30/150/300 mM		
<b>Buffer detergent</b>	SDS 0.2mM & 0.1%, Tween 20 0.1%, TritonX 0.01%		
<b>Grids</b>	Quantifoil (Cu, Au, C-Flat) & Lacey	Fail	<b>Grids Nature</b>
	UTC Quantifoil	<b>Success</b>	
	UltrAufoil	Partial success	

**Table 2.** Review of all grids and parameters screened and the impact on particles distribution.

Grid type	OD <sub>280nm</sub> [mg/ml] ( $\mu$ M)	Microsc ope	Data collection parameters				Med. CTF	Ptcs. nbrs.	Map. Res.	b-factor	
<b>UTC Quan. 1.2/1.3<math>\mu</math>m</b>	0.15 [0,12] (0.73)	Glacios	8787 mic.	240k mag.	Falcon 4	0.58 A/pix	60e/A <sup>2</sup>	3.2Å	711 008	3.1Å	130.2
UTC Quan. 1.2/1.3 $\mu$ m	0.34 [0,27] (1,6)	Glacios	1000 mic.	240k mag.	Falcon 4	0.58 A/pix	60 e/A <sup>2</sup>	3.7Å	70 071	5.97Å	374.4
<b>Quantifoil 300 1.2/1.3<math>\mu</math>m</b>	0.3 [0,23] (1,45)	Glacios	8870 mic.	240k mag.	Falcon 4i	0.60 A/pix	50e/A <sup>2</sup>	3.5Å & 8.5Å	489 795	<6Å	331
Quantifoil 300 1.2/1.3 $\mu$ m	0.1 [0,08] (0,48)	Glacios	1448 mic.	120k mag.	falcon 3	1.12 A/pix	60e/A <sup>2</sup>	7.5Å	***	***	***
<b>UltrAufoil 300 0.6/1 <math>\mu</math>m</b>	0.1 [0,08] (0,48)	Titan Krios	4231 mic.	102K mag.	Falcon 3i	0.86 A/pix	60e/A <sup>2</sup>	3.8Å	740 544	6<Å<7.2	741.3
UltrAufoil 300 R1.2/1.3 $\mu$ m	0.1 [0,08] (0,48)	Glacios	1861 mic.	120k mag	Falcon3	1.12 A/pix	60e/A <sup>2</sup>	7.5Å	91 501	7.4Å	662.8

**Table 3.** Summary of data collection parameters and data processing results. Only the grids in bold are referenced in this article. The fourth column shows the number of micrographs (mic.) remaining after screening (usually by CTF fit resolution, but also by visual inspection). The fifth column gives the magnification used during data collection. The sixth column gives the pixel size of the RAW movies. The ninth column gives the median CTF value of the whole dataset and was determined visually according to the plot  $CTF_{fit_{resolution}}=f(\text{defocus})$  made in cryoSPARC (two values are given for a bimodal distribution). B-factor is given only for the global refinement.

Interfaces sHER2-Fabs	
sHER2	Fab
sHER2-trastuzumab interface	
(1) Asp560, Gln561, Pro557	<b>CDR H2</b> Arg50, Tyr57, Arg59
(2) Asp570, Pro572, Phe573, Lys593	<b>CDR H1</b> Tyr33, & <b>CDR H3</b> Trp99, Tyr105
(3) Trp592, Pro603 Glu602	<b>CDR H3</b> Phe104, Ala106, <b>CDR L2</b> Leu46, Tyr49, Tyr55, Phe53 <b>CDR L1</b> Asp30
(4) Asp585	<b>CDR H3</b> , Arg98
sHER2-pertuzumab interface	
(1) Leu244, His245	<b>CDR H2</b> Gly55
(2) Tyr252, Phe257,	<b>CDR H2</b> Lys58, Tyr59, Gln61 & <b>CDR L3</b>

Thr254, Asp255	Tyr94
(3) Val286, Thr290, Pro294, Leu295	<b>CDR H1</b> Thr33, Tyr32 & <b>CDR H3</b> Pro98, Leu96
(4) Ser288, Thr290	<b>CDR H1</b> Asp31, <b>CDR H2</b> Asp 53
(5) His286	<b>CDR H3</b> Tyr99B
(6) Pro315	<b>CDR L2</b> Tyr49, Tyr53

**sHER2 domains & Fab CDRs**

<b>Domain I:</b> 1-185	<b>Domain III:</b> 320-485	<b>Pertuzumab</b> : CDR L1 (K24-A34); CDR L2 (S50-Y55); CDR L3 (Q89-T97); CDR H1 (G26-D35), CDR H2 (A49-G65), CDR H3 (A93-Y103)
<b>Domain II:</b> 186-219	<b>Domain IV:</b> 486-630	<b>Trastuzumab:</b> CDR L1 (R24-A34), CDR L2 (S50-Y55), CDR L3 (Q89-T97), CDR H1 (G26-H35), CDR H2 (R50-K65), CDR H3 (S97-Y109)

**Table 4.** Top, summary of the different epitope/paratope interfaces defined by the amino acids involved in the interactions. Bottom, definition of the HER2 domains and CDR of the two Fabs. CDR are defined according to *Ling et al.* (Ling et al. 2018)

**High resolution maps quoted in the text**

		<b>GS-FSC resolution</b>
<b>Global refinement (after CTFrefine)</b>	Global homogeneous refinement map	3.30Å
<b>Local refinement (after CTFrefine)</b>	HER2-pertuzumab & HER2-trastuzumab	3.05Å & 3.33Å

<b>Composite map (combined using phenix)</b>	Composite map	3.19Å
<b>3DFlex 711k ptcs (particles after ctf refinement)</b>	3Dflex-711k	3.54Å
<b>3DFlex 286k ptcs</b>	3DFlex-286k	3.5Å

**Table 5.** Summary of the maps generated from the dataset collected on the UTC-Quantifoil grid and presented in the article.

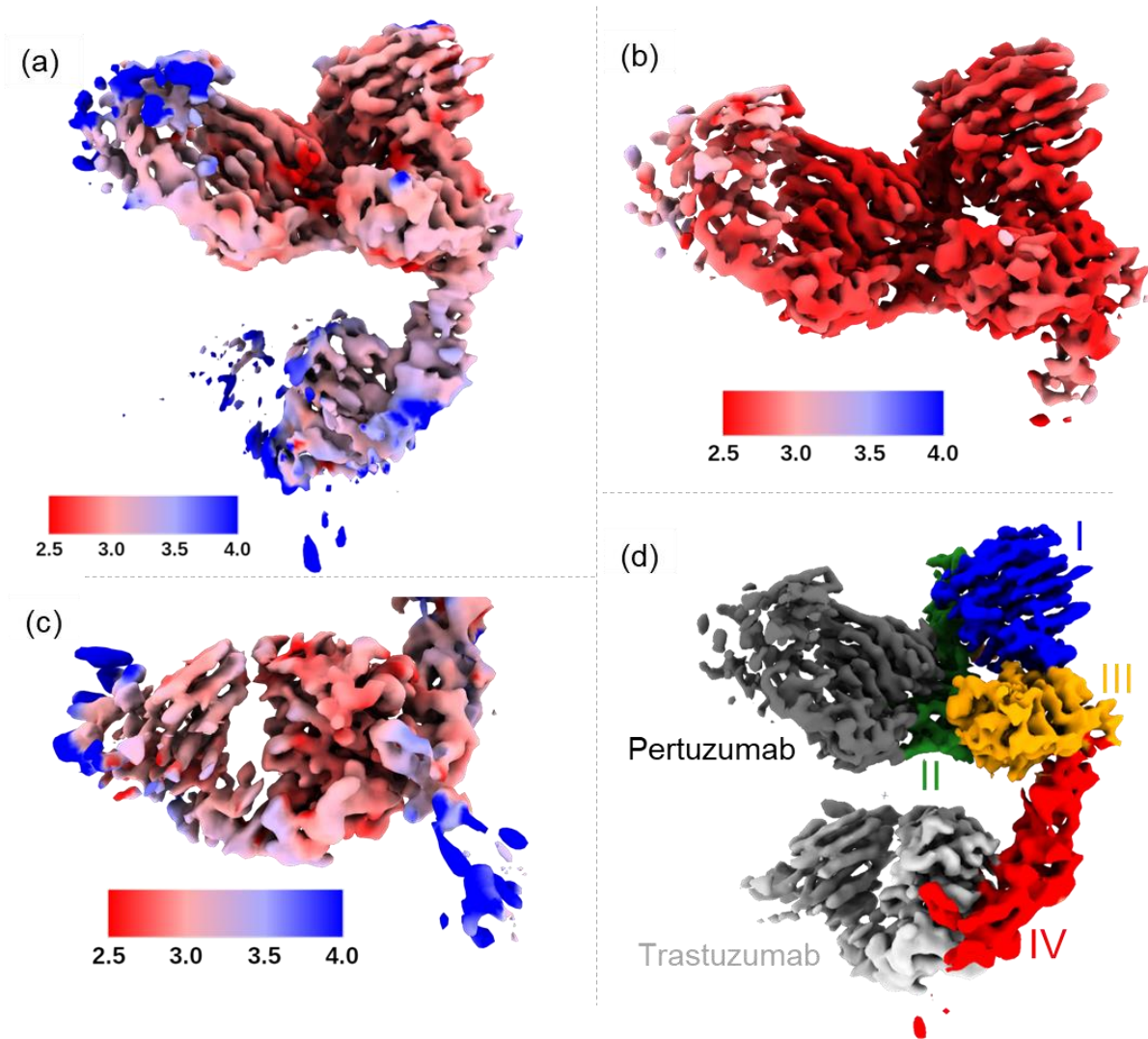
<b>Model</b>		
<b>Composition (#)</b>	<b>Composite map (8PWH, EMD-17993)</b>	<b>Global map (8G6J, EMD-18188)</b>
Chains	5	5
Atoms	11535 (Hydrogens: 0)	11535 (Hydrogens: 0)
Residues	Protein: 1494 Nucleotide: 0	Protein: 1494 Nucleotide: 0
Water	0	0
Ligands	BMA: 2 NAG: 7	BMA: 2 NAG: 7
<b>Bonds (RMSD)</b>		
Length (Å) (# > 4σ)	0.004 (0)	0.007 (0)
Angles (°) (# > 4σ)	0.612 (2)	0.735 (1)
MolProbity score	2.65	2.90
Clash score	15.36	19.52
Ramachandran plot (%)		
Outliers	0.07	0.07
Allowed	7.55	9.70
Favored	92.39	90.23
<b>Rama-Z (Ramachandran plot Z-score, RMSD)</b>		
whole (N = 1484)	-1.53 (0.22)	-1.61 (0.22)
helix (N = 52)	-1.67 (0.67)	-1.45 (0.70)
sheet (N = 434)	-0.53 (0.25)	-0.59 (0.26)
loop (N = 998)	-1.25 (0.20)	-1.31 (0.20)



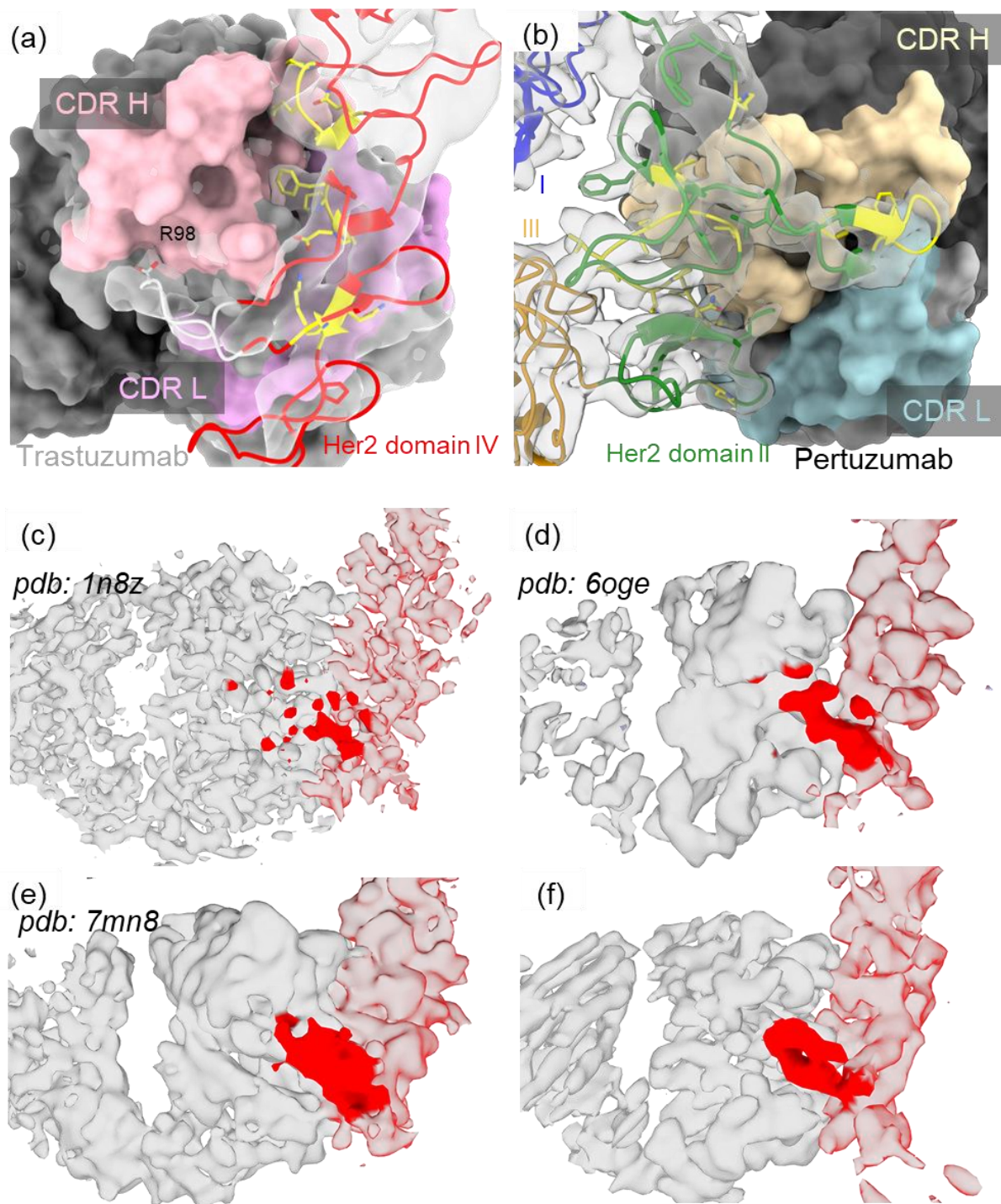
Rotamer outliers (%)	4.36	5.69
C $\beta$ outliers (%)	0.00	0.00
<b>Peptide plane (%)</b>		
Cis proline/general	8.7/0.0	8.7/0.0
Twisted proline/general	0.0/0.0	0.0/0.0
CaBLAM outliers (%)	4.48	3.93
<b>ADP (B-factors)</b>		
Iso/Aniso (#)	11535/0	11535/0
<b>min/max/mean</b>		
Protein	48.43/278.76/138.51	53.07/1018.44/225.27
Nucleotide	N.A.	N.A.
Ligand	50.89/184.43/126.51	114.54/263.93/185.13
Water	N.A.	N.A.
<b>Occupancy</b>		
Mean	1.00	1.00
occ = 1 (%)	100.00	100.00
0 < occ < 1 (%)	0.00	0.00
occ > 1 (%)	0.00	0.00
<b>Data</b>		
<b>Box</b>		
Lengths (Å)	105.56, 131.08, 148.48	105.56, 129.92, 148.48
Angles (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Supplied Resolution (Å)	3,1	3.1
Resolution Estimates (Å)	Masked : Unmasked	Masked : Unmasked
d FSC (half maps; 0.143)	N.A.	N.A.
d 99 (full/half1/half2)	4.0/N.A./N.A. : 4.0/N.A./N.A.	4.2/N.A./N.A. : 4.2/N.A./N.A.
d model	3.7 : 3.6	3.8 : 3.8
d FSC model (0/0.143/0.5)	2.9/3.1/3.6 : 3.0/3.2/3.7	3.2/3.4/4.0 : 3.3/3.5/4.1
Map min/max/mean	-34.79/63.03/0.71	-0.44/1.02/0.01
<b>Model vs. Data</b>		
CC (mask)	0.76	0.75
CC (box)	0.84	0.86
CC (peaks)	0.70	0.70
CC (volume)	0.76	0.75
Mean CC for ligands	0.59	0.66

**Table 6. Models review**, refinement statistics and cross-correlation information obtained using the composite map or the global map.

## Main figures

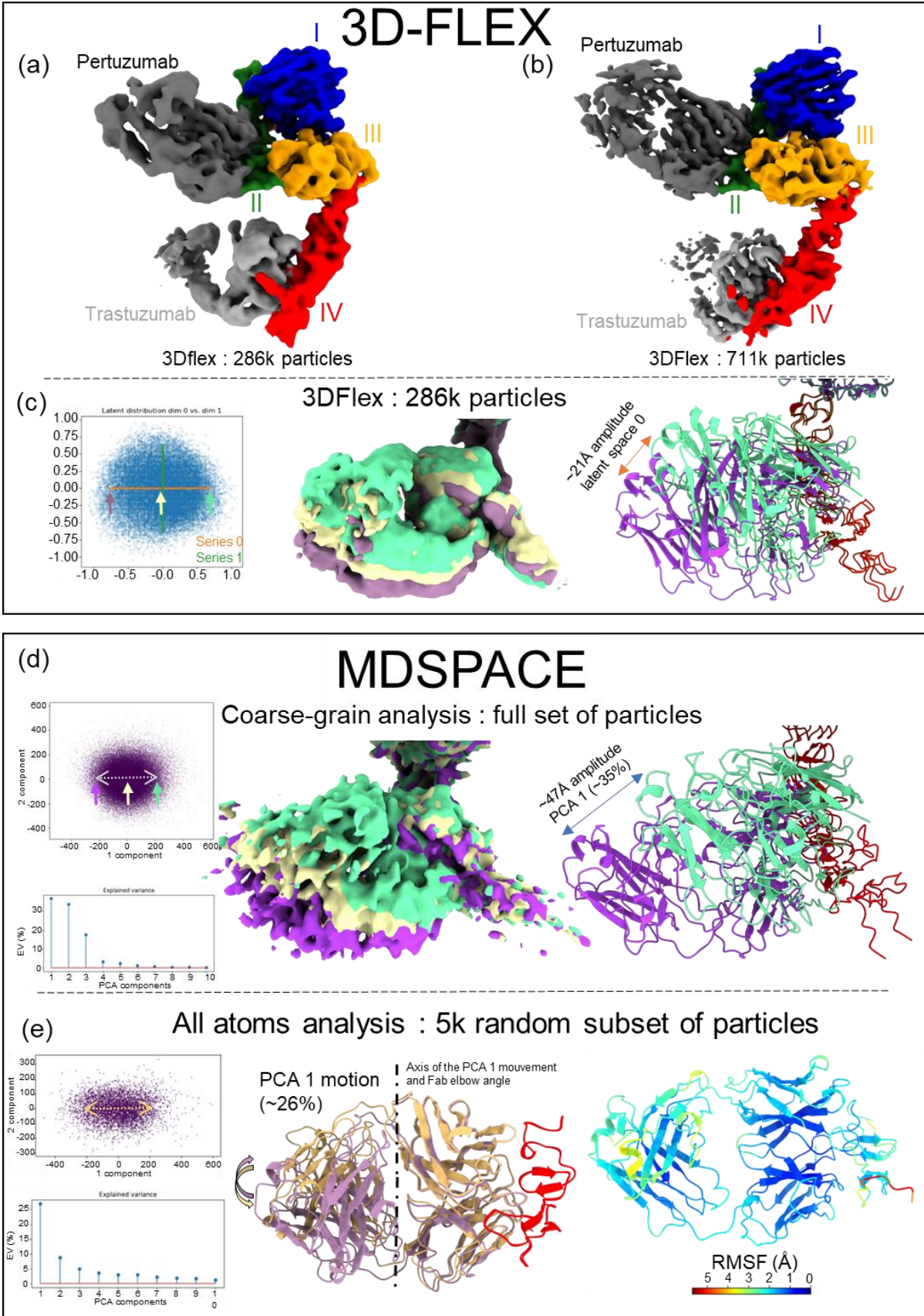


**Figure 1: Baseline protocol vs focused and composite maps for the HTP complex.** Panel (a) shows the global map obtained from classic homogeneous refinement (baseline protocol). (b) and (c) maps are those obtained from focused refinements of the pertuzumab (b) and trastuzumab (c) regions. (d), composite map combining the high resolution of the two local refinement maps (b,c) in one unique map. (a), (b), (c) are colored according to local resolution. (d) sHER2 is colored by domains (blue: I; green: II; yellow: III; red: IV), the pertuzumab Fab is colored in dark grey and the trastuzumab Fab in light grey.



**Figure 2: Pertuzumab and trastuzumab interfaces and overview of previously published maps.** Panel **a** shows the interface between HER2 and trastuzumab. Fab is shown as surface in grey (heavy chain) and light grey (light chain) with CDR's in pink (CDR H1-3) and purple (CDR L1-3). Experimental cryo-EM map of sHER2 is shown in transparent white. The four loops of the HER2 epitopes are in yellow for the three (557-561, 570-573, and 602-603) previously established (Cho et al., 2003) and in white for the fourth (581-593, this work). In red is the sHER2 domain IV. **b** shows the pertuzumab region. Fab is shown as surface in black (heavy chain) and light black (light chain) with CDR's in pale gold (CDR H1-3) and pale blue (CDR L1-3). Experimental cryo-EM map is shown in transparent white. sHER2 chain is colored by domains (blue for domain I, orange for domain II, and green for domain III). **c** is the x-ray map of 1n8z (Hao et al. 2019), **d** the cryo-EM map of 6oge (Hao et al. 2019), **e** the cryo-EM map of 7mn8 (HER2-HER3-NRG1 $\beta$ -trastuzumab complex) (Diwanji et al. 2021) and **f** the focused cryo-EM map of Figure 1c (this work). The loop 581-590

*that we propose to be considered as part of the trastuzumab epitope is show in red in all maps, while sHER2 is in light red and Trastuzumab is colorless.*

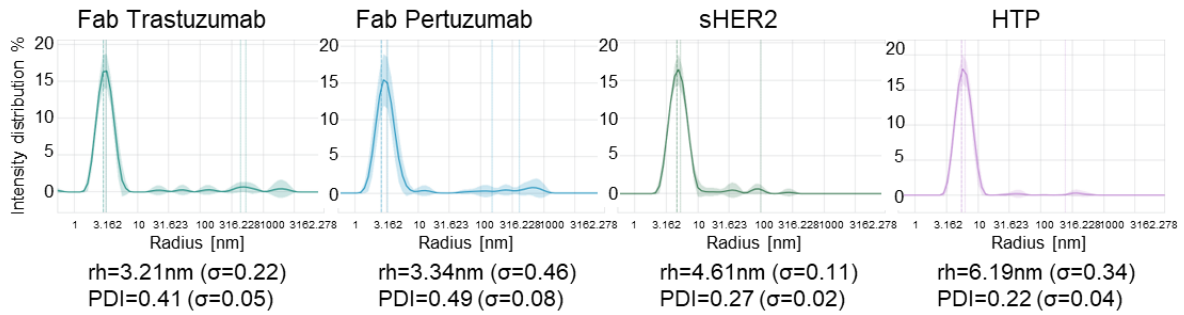


**Figure 3: Continuous conformational variability analysis.** (a) shows the 3D flex map obtained with the 286k subset of particles after RSPFilter screening, and (b) shows the 3DFlex map obtained using the

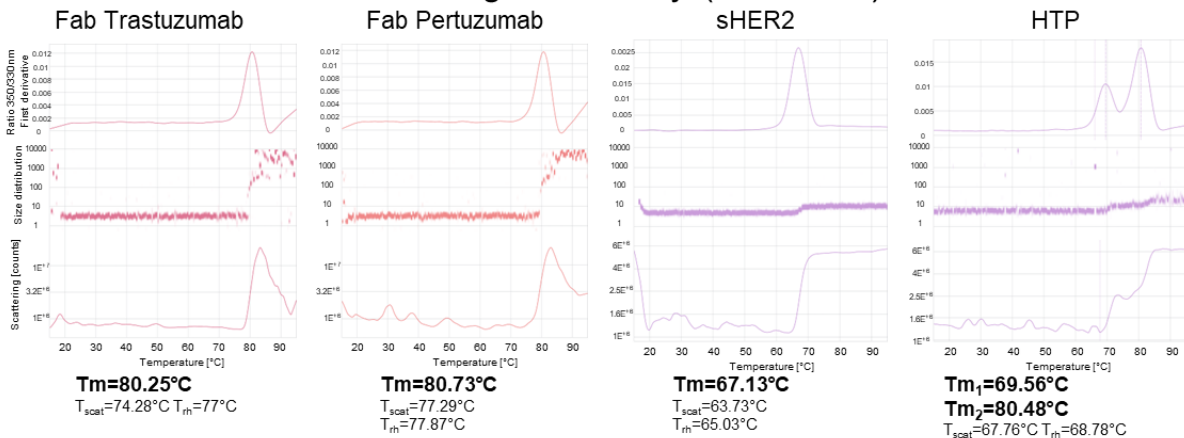
full dataset of particles. Both are colored by domain (I=Blue, II=Green, III=Orange, IV= Red). **(c)** shows the 286k subset 3Dflex conformational landscape visualization (left, orange line) resulting in an amplitude of motion of 21Å in the trastuzumab region. Note the loop density staying visible and moving collectively with the trastuzumab (see also the supplementary Movies). **(d), (e) MDSPACE analysis. (d)**, analysis of the major motions using the full set of particles and a coarse-grained (C $\alpha$ ) HTP model. Left, principal component analysis: Bottom, variance explained by each principal component. Top, two-dimensional plot of the first two principal components. Right, 3D reconstructions and coarse-grained models: The fitted models and their corresponding maps for the first principal component extreme values after alignment to the pertuzumab region are shown, as well as one intermediate map (pale gold). Note that a large amplitude of motion of 47Å for trastuzumab+HER2 IV is found. **(e)**, local motion of the trastuzumab region, analyzed with a subset of 5000 particles and all-atom simulations with subsequent superposition on trastuzumab. Left, principal component analysis: Bottom, variance explained by each principal component. Top, two-dimensional plot of the first two principal components. Middle, two models showing the two extremes for the first principal component (26%) after alignment to the domain IV of sHER2. It shows the main motion is between the constant and variable parts of trastuzumab in accordance with the characteristic elbow angle of the fab. Right panel shows the root mean square fluctuation among the 5000 structures.

## Supplementary figures

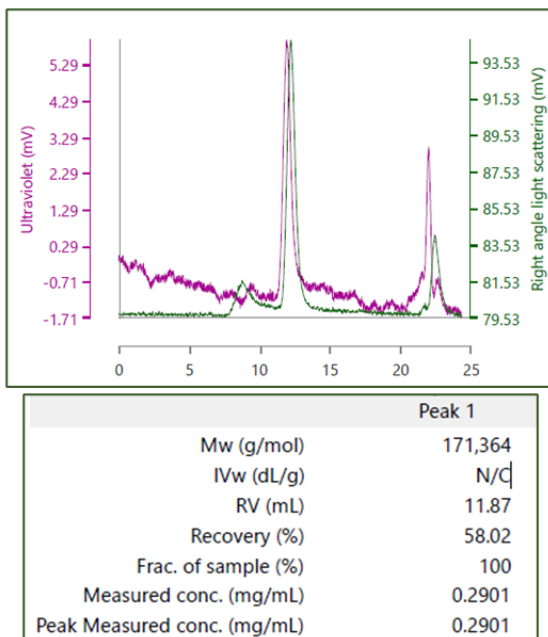
### Dynamic Light Scattering (DLS)



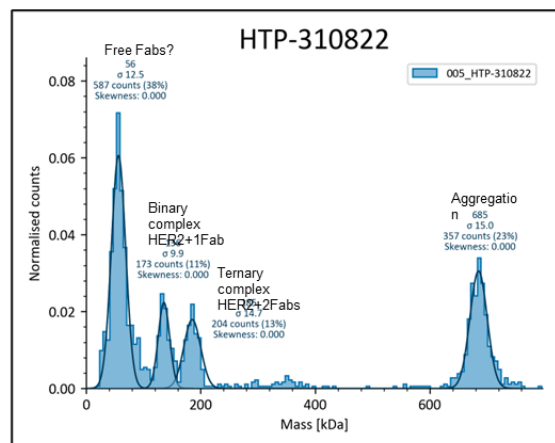
### Nano-Differential Scanning Fluorimetry (nano-DSF)



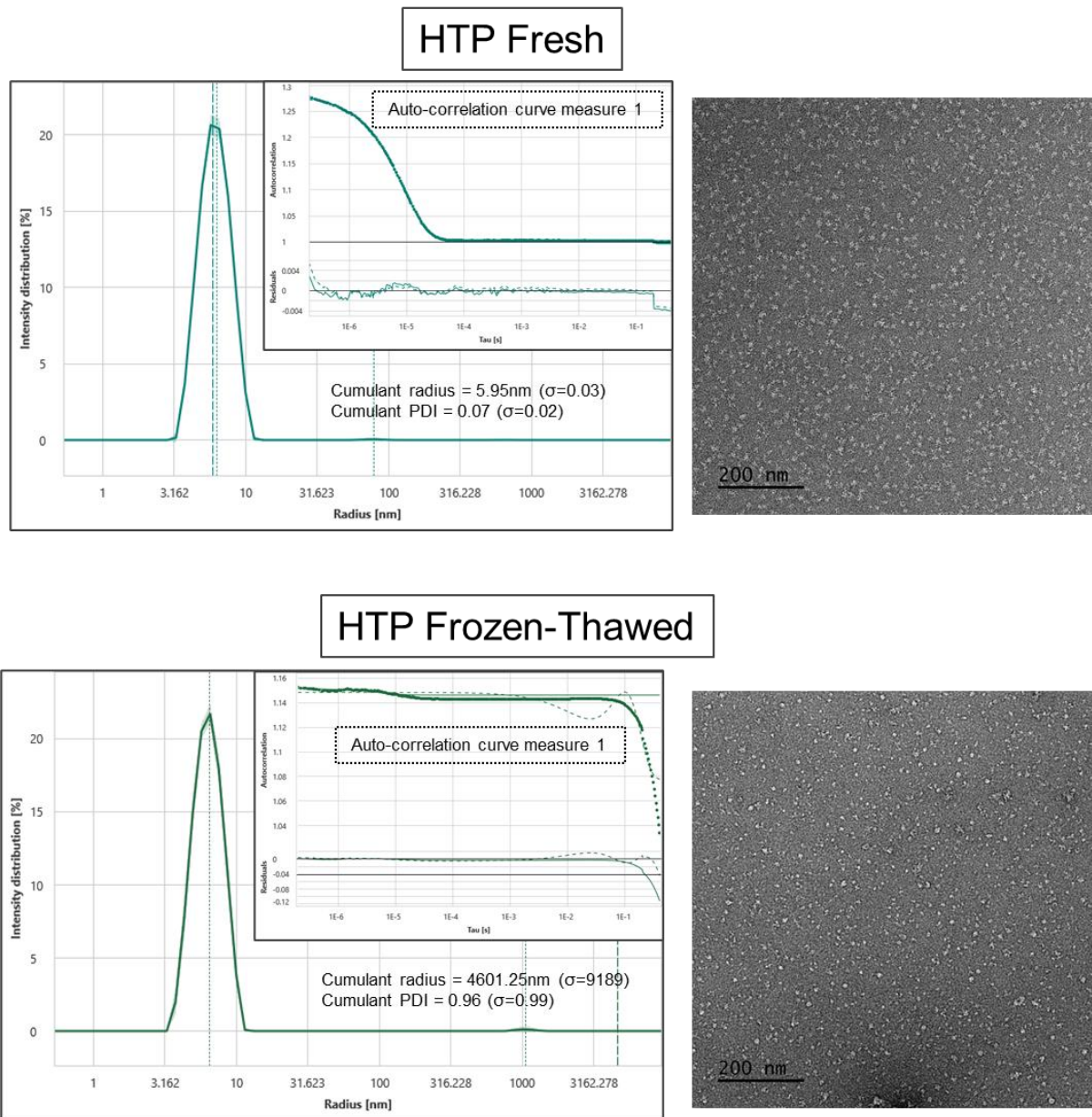
### SEC-SLS



### Mass-Photometry

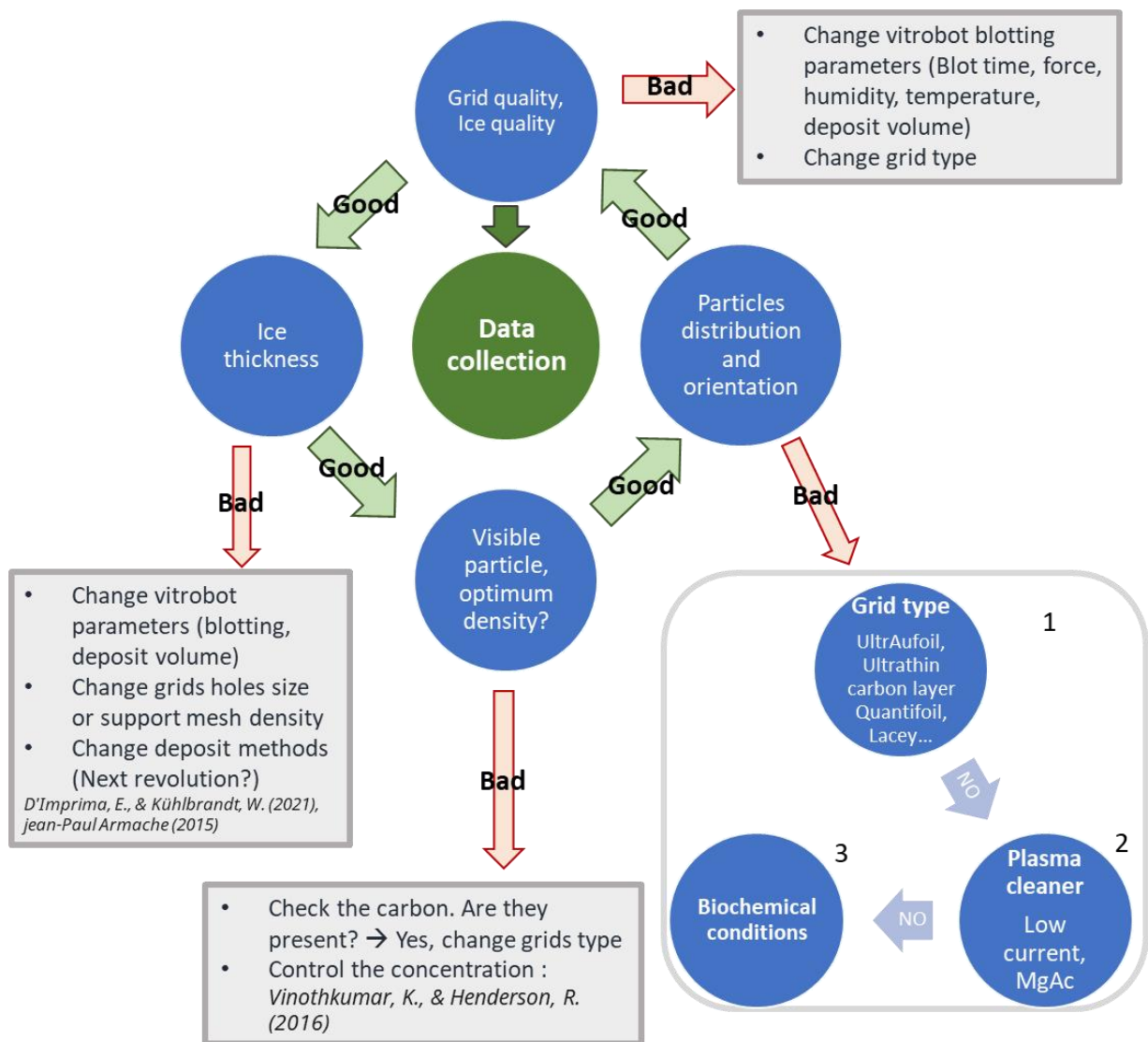
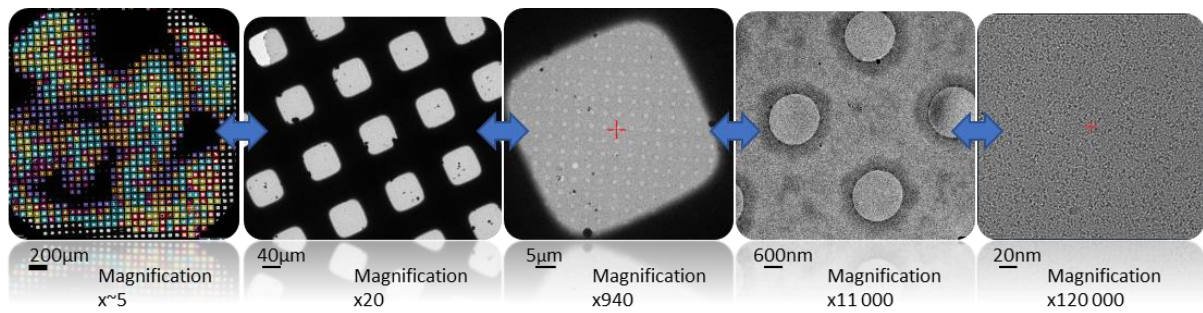


**Supplementary figure 1. Biophysical sample characterization.** DLS, nano-DSF, SEC-SLS and Mass-photometry measures. DLS and Nano-DSF show typical results for each Fabs, sHER2 and the HTP complex. The SEC-SLS panel shows the result of the SLS using HTP. Mass-photometry panel shows the measure on the flash-frozen/thawed HTP sample.



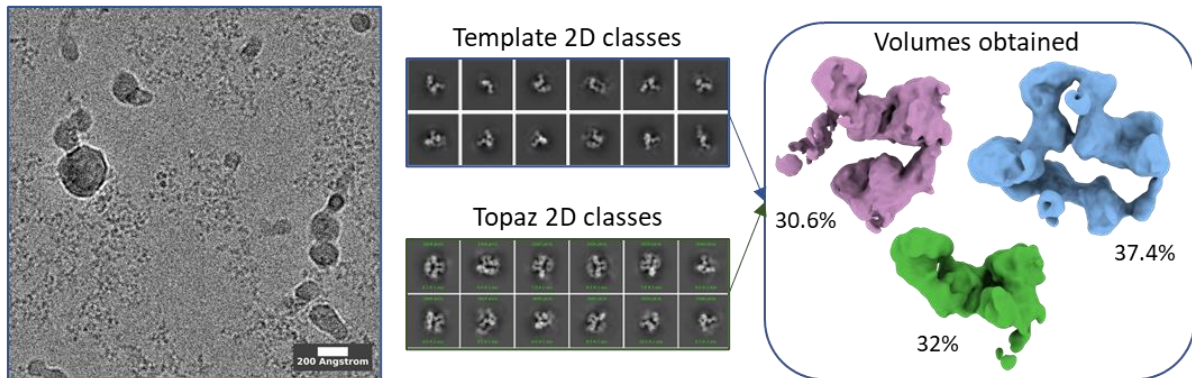
**Supplementary figure 2.** Sample DLS measurement (nanotemper panta) and negative staining comparison of same sample before and after flash-frozen/thawed cycle.





**Supplementary figure 3.** Synthetic pipeline for grids screening and decision tree for troubleshooting.

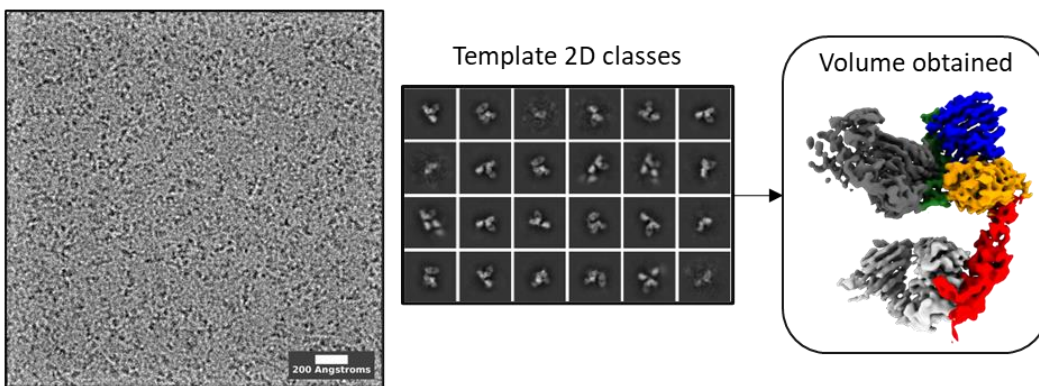
- Quantifoil (200kV)



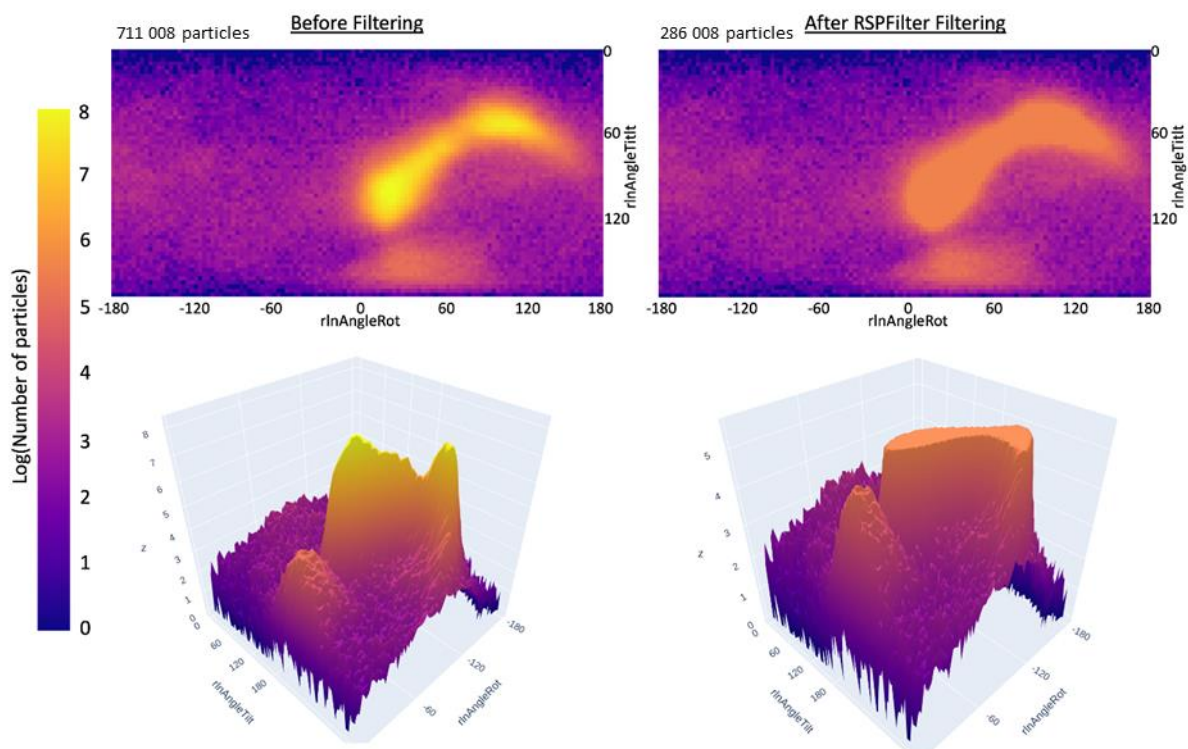
- UltrAuFoil (300k)



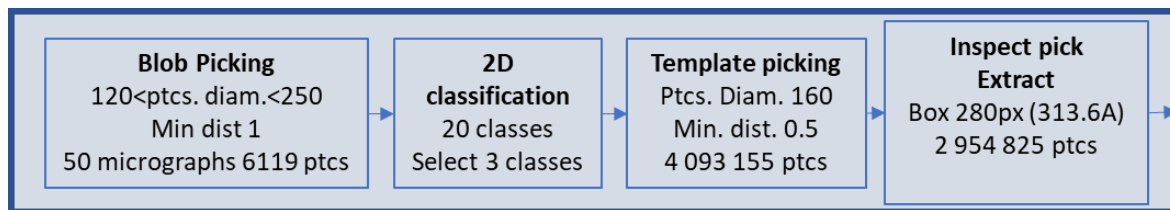
- UTC Quantifoil (200kV)



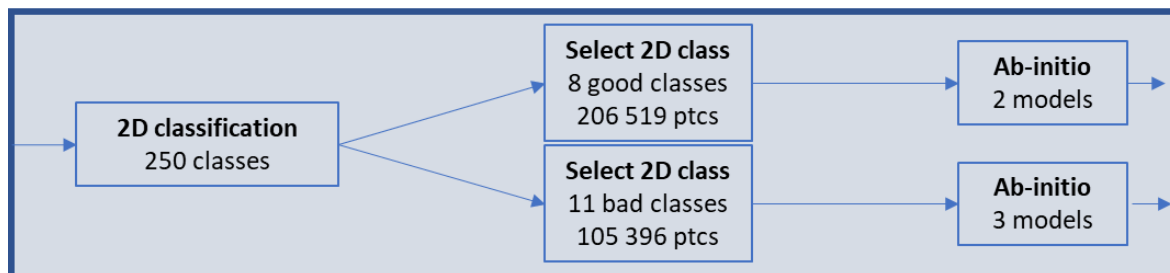
**Supplementary figure 4.** To the left is shown typical micrograph aspect, in the middle few representatives 2D classes and to the right the volumes obtained according to the grids used. Upper volumes were obtained when using QF 1.2/1.3 $\mu$ m grids. The violet is the classic HTP trimer. However, it is worth to note that 3D classification with particles of this class reveal a strong heterogeneity (not shown). The doublet is in blue and the partial complex (Pertuzumab/sHER2) is in green. Middle panel correspond to data obtained when using UF 0.6/1 $\mu$ m, two volumes can be unambiguously identified as the HTP in violet and the partial complex. Two others are more difficult to assign but could arise from strongly deteriorated particles such as Fab and trastuzumab/sHER2 or other. Lower panel correspond to data obtained when using UTC quantifoil 1.2/1.3 $\mu$ m. Sample appear highly homogeneous, and a high-resolution map could be obtained without any trace of dissociated complex.



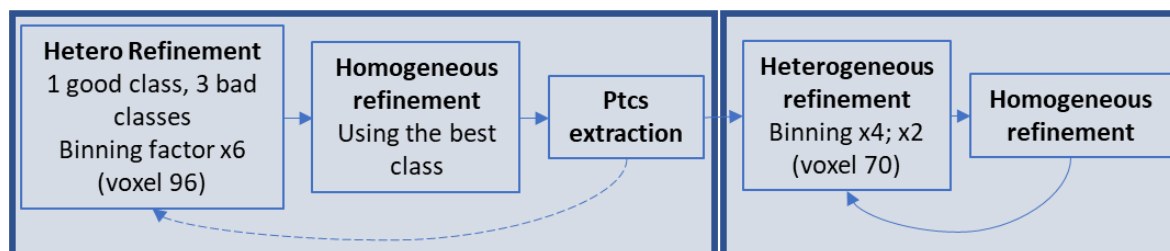
**Supplementary figure 5. Particle angular distribution from the global refinement before and after RSPFilter filtering.** Left is the angular distribution 2D plot of the dataset before RSPFilter clean-up and right is after clean-up. Lower panels show the same plots visualized in 3D. The scale bar color is the same for all plots (2D and 3D). Only the particles from the more represented views are removed without impacting the rare views of the object.



Particles « picking »

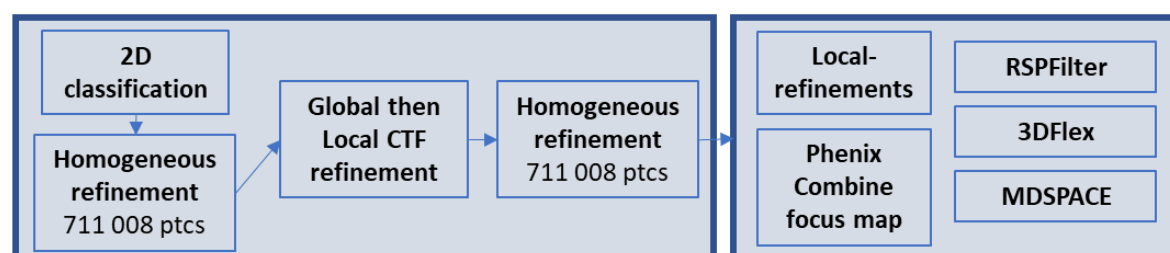


3D classification volumes preparation



Particles enrichment

Particles screening



Further processing

**Supplementary figure 6.** Overview of the data processing strategy used for UTC-Quantifoil grid.

**Supplementary data movie 1.** Movie showing the 3DFlex movement described by the latent space 0 using the full set of 711k particles.

**Supplementary data movie 2.** Movie showing the 3DFlex movement described by the latent space 0 using the sub-set of 286k particles.

**Supplementary data movie 3.** Movie showing the MDSPACE movement described by the principal component 1 from the coarse grain analysis using the full set of 711k particles.

- Asarnow, D., Palovcak, E. & Cheng, & Y. (2019). *Zenodo*  
<https://doi.org/10.5281/zenodo.3576630>.
- Bai, X., Sun, P., Wang, X., Long, C., Liao, S., Dang, S., Zhuang, S., Du, Y., Zhang, X., Li, N., He, K. & Zhang, Z. (2023). *Cell Discov* **9**, 18.
- Ben-Kasus, T., Schechter, B., Lavi, S., Yarden, Y. & Sela, M. (2009). *Proc. Natl. Acad. Sci.* **106**, 3294–3299.
- Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A. J. & Berger, B. (2019). *Nat Methods* **16**, 1153–1160.
- Berrow, N., Marco, A. de, Lebendiker, M., Garcia-Alai, M., Knauer, S. H., Lopez-Mendez, B., Matagne, A., Parret, A., Remans, K., Uebel, S. & Raynal, B. (2021). *Eur Biophys J* **50**, 453–460.
- Blumenthal, G. M., Scher, N. S., Cortazar, P., Chattopadhyay, S., Tang, S., Song, P., Liu, Q., Ringgold, K., Pilaro, A. M., Tilley, A., King, K. E., Graham, L., Rellahan, B. L., Weinberg, W. C., Chi, B., Thomas, C., Hughes, P., Ibrahim, A., Justice, R. & Pazdur, R. (2013). *Clin Cancer Res* **19**, 4911–4916.
- Brown, H. G. & Hanssen, E. (2022). *Commun Biology* **5**, 817.
- Cho, H.-S., Mason, K., Ramyar, K. X., Stanley, A. M., Gabelli, S. B., Denney, D. W. & Leahy, D. J. (2003). *Nature* **421**, 756–760.
- Croll, T. I. (2018). *Acta Crystallogr Sect D Struct Biology* **74**, 519–530.
- Diwanji, D., Trenker, R., Thaker, T. M., Wang, F., Agard, D. A., Verba, K. A. & Jura, N. (2021). *Nature* **600**, 339–343.
- EMA, E. M. A. (2000). <https://www.ema.europa.eu/en/medicines/human/EPAR/herceptin>.
- EMA, E. M. A. (2013). <https://www.ema.europa.eu/en/medicines/human/EPAR/perjeta>.
- FDA (1998). <https://www.drugs.com/history/herceptin.html>.
- FDA, F. and D. A. (2012). <https://www.drugs.com/perjeta.html>.
- Franklin, M. C., Carey, K. D., Vajdos, F. F., Leahy, D. J., Vos, A. M. de & Sliwkowski, M. X. (2004). *Cancer Cell* **5**, 317–328.
- Fu, W., Wang, Y., Zhang, Y., Xiong, L., Takeda, H., Ding, L., Xu, Q., He, L., Tan, W., Bethune, A. N. & Zhou, L. (2014). *Mabs* **6**, 978–990.

- Garrett, T. P. J., McKern, N. M., Lou, M., Elleman, T. C., Adams, T. E., Lovrecz, G. O., Zhu, H.-J., Walker, F., Frenkel, M. J., Hoyne, P. A., Jorissen, R. N., Nice, E. C., Burgess, A. W. & Ward, C. W. (2002). *Cell* **110**, 763–773.
- Ghosh, R., Narasanna, A., Wang, S. E., Liu, S., Chakrabarty, A., Balko, J. M., González-Angulo, A. M., Mills, G. B., Penuel, E., Winslow, J., Sperinde, J., Dua, R., Pidaparathi, S., Mukherjee, A., Leitzel, K., Kostler, W. J., Lipton, A., Bates, M. & Arteaga, C. L. (2011). *Cancer Res* **71**, 1871–1882.
- Graus-Porta, D., Beerli, R. R., Daly, J. M. & Hynes, N. E. (1997). *Embo J* **16**, 1647–1655.
- Hao, Y., Yu, X., Bai, Y., McBride, H. J. & Huang, X. (2019). *Plos One* **14**, e0216095.
- Harastani, M., Vuillemot, R., Hamitouche, I., Moghadam, N. B. & Jonic, S. (2022). *J Struct Biol* **214**, 107906.
- Howie, L. J., Scher, N. S., Amiri-Kordestani, L., Zhang, L., King-Kallimanis, B. L., Choudhry, Y., Schroeder, J., Goldberg, K. B., Kluetz, P. G., Ibrahim, A., Sridhara, R., Blumenthal, G. M., Pazdur, R. & Beaver, J. A. (2019). *Clin Cancer Res* **25**, 2949–2955.
- Karunakaran, D., Tzahar, E., Beerli, R. R., Chen, X., Graus-Porta, D., Ratzkin, B. J., Seger, R., Hynes, N. E. & Yarden, Y. (1996). *Embo J* **15**, 254–264.
- Kobayashi, C., Jung, J., Matsunaga, Y., Mori, T., Ando, T., Tamura, K., Kamiya, M. & Sugita, Y. (2017). *J. Comput. Chem.* **38**, 2193–2206.
- Ling, W.-L., Lua, W.-H., Poh, J.-J., Yeo, J. Y., Lane, D. P. & Gan, S. K.-E. (2018). *Front Immunol* **9**, 469.
- Liu, X., Fang, Y., Li, Y., Li, Y., Qi, L. & Wang, X. (2022). *Frontiers Oncol* **12**, 894861.
- Molina, M. A., Codony-Servat, J., Albanell, J., Rojo, F., Arribas, J. & Baselga, J. (2001). *Cancer Res.* **61**, 4744–4749.
- Nahta, R., Hung, M.-C. & Esteva, F. J. (2004). *Cancer Res* **64**, 2343–2346.
- Nami, B., Maadi, H. & Wang, Z. (2018). *Cancers* **10**, 342.
- Olayioye, M. A., Neve, R. M., Lane, H. A. & Hynes, N. E. (2000). *Embo J* **19**, 3159–3167.
- Passmore, L. A. & Russo, C. J. (2016). *Methods Enzymol* **579**, 51–86.
- Peckys, D. B., Korf, U. & Jonge, N. de (2015). *Sci Adv* **1**, e1500165.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H. & Ferrin, T. E. (2021). *Protein Sci* **30**, 70–82.
- Plotly (2015). Collaborative data science.
- Pohlmann, P. R., Mayer, I. A. & Mernaugh, R. (2009). *Clin Cancer Res* **15**, 7479–7491.

- Punjani, A. & Fleet, D. J. (2021). *Biorxiv* 2021.04.22.440893.
- Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. (2017). *Nat Methods* **14**, 290–296.
- Raynal, B., Brûlé, S., Uebel, S. & Knauer, S. H. (2021). *Methods Mol Biology* **2263**, 3–46.
- Rice, W. J., Cheng, A., Noble, A. J., Eng, E. T., Kim, L. Y., Carragher, B. & Potter, C. S. (2018). *J Struct Biol* **204**, 38–44.
- Richard, S., Selle, F., Lotz, J.-P., Khalil, A., Gligorov, J. & Soares, D. G. (2016). *An. Da Acad. Bras. Ciências* **88**, 565–577.
- Rosa-Trevín, J. M. de la, Quintana, A., Cano, L. del, Zaldívar, A., Foche, I., Gutiérrez, J., Gómez-Blanco, J., Burguet-Castell, J., Cuenca-Alba, J., Abrishami, V., Vargas, J., Otón, J., Sharov, G., Vilas, J. L., Navas, J., Conesa, P., Kazemi, M., Marabini, R., Sorzano, C. O. S. & Carazo, J. M. (2016). *J Struct Biol* **195**, 93–99.
- Scheuer, W., Friess, T., Burtscher, H., Bossenmaier, B., Endl, J. & Hasmann, M. (2009). *Cancer Res* **69**, 9330–9336.
- Sonn-Segev, A., Belacic, K., Bodrug, T., Young, G., VanderLinden, R. T., Schulman, B. A., Schimpf, J., Friedrich, T., Dip, P. V., Schwartz, T. U., Bauer, B., Peters, J.-M., Struwe, W. B., Benesch, J. L. P., Brown, N. G., Haselbach, D. & Kukura, P. (2020). *Nat Commun* **11**, 1772.
- Tubiana, T. (2022). *Zenodo* <https://doi.org/10.5281/zenodo.7413914>.
- Vinothkumar, K. R. & Henderson, R. (2016). *Q Rev Biophys* **49**, e13.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Walt, S. J. van der, Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., Mulbregt, P. van, Contributors, S. 10, Vijaykumar, A., Bardelli, A. P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C. N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D. A., Hagen, D. R., Pasechnik, D. V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G. A., Ingold, G.-L., Allen, G. E., Lee, G. R., Audren, H., Probst, I., Dietrich, J. P., Silterra, J., Webber, J. T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J. L., Cardoso, J. V. de M., Reimer, J., Harrington, J., Rodríguez, J. L. C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N. J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P. A., Lee, P., McGibbon, R. T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T. J., Robitaille, T. P., Spura, T., Jones, T. R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y. O. & Vázquez-Baeza, Y. (2020). *Nat Methods* **17**, 261–272.

- Vuillemot, R., Mirzaei, A., Harastani, M., Hamitouche, I., Fréchin, L., Klaholz, B. P., Miyashita, O., Tama, F., Rouiller, I. & Jonic, S. (2023). *J Mol Biol* 167951.
- Vuillemot, R., Miyashita, O., Tama, F., Rouiller, I. & Jonic, S. (2022). *J Mol Biol* **434**, 167483.
- Weissenberger, G., Henderikx, R. J. M. & Peters, P. J. (2021). *Nat Methods* **18**, 463–471.