# Dynamical rearrangement of human epidermal growth factor receptor 2 upon antibody binding: effects on the dimerization

Pedro R. Magalhães[1], Miguel Machuqueiro[1], José G. Almeida[2], André Melo [3], M. Natalia D. S. Cordeiro[3], Sandra Cabo Verde [4], Zeynep H. Gümüş [5], Irina S. Moreira [2], João D. G. Correia [4]*, Rita Melo[4,2]*

[1] Centro de Química e Bioquímica and Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal

[2] CNC - Center for Neuroscience and Cell Biology; Rua Larga, FMUC, Polo I, 1ºandar, Universidade de Coimbra, 3004-517; Coimbra, Portugal

[3] REQUIMTE/LAQV, Faculdade de Ciências da Universidade do Porto, Departamento de Química e Bioquímica, Rua do Campo Alegre, 4169-007 Porto, Portugal

[4] Centro de Ciências e Tecnologias Nucleares e Departamento de Engenharia e Ciências Nucleares, Instituto Superior Técnico, Universidade de Lisboa, CTN, Estrada Nacional 10 (km 139,7), 2695-066 Bobadela LRS, Portugal

[5] Department of Genetics and Genomics and Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

Corresponding authors.

E-mail addresses: jgalamba@ctn.tecnico.ulisboa.pt (J. D. G. Correia); ritamelo@ctn.tecnico.ulisboa.pt (R. Melo).

# Abstract

Human epidermal growth factor 2 (HER2) is a ligand-free tyrosine kinase receptor of the HER family that is overexpressed in some of the most aggressive tumours. Treatment of HER2+ breast cancers with the humanized monoclonal anti-HER2 antibody (Trastuzumab) revealed highly effective, encouraging the development of various HER2-specific antibodies, kinase inhibitors and dimerization inhibitors for cancer therapy. Although it is known that HER2 dimerization involves a specific region of its extracellular domain, the so-called "dimerization arm", the mechanism of dimerization inhibition remains uncertain. However, uncovering how antibody interactions lead to inhibition of HER2 dimerization is of key importance in understanding its role in tumour progression and therapy. Herein, we employed several computational modelling techniques for a molecular-level understanding of the interactions between HER and specific anti-HER2 antibodies, namely an antigen-binding (Fab) fragment (F0178) and a single chain variable fragment from Trastuzumab (scFv). Specifically, we investigated the effects of antibody-HER2 interactions on the key residues of "dimerization arm" from molecular dynamics (MD) simulations of unbound HER (in a total of 1 µs), as well as scFv:HER2 and F0178:HER2 complexes (for a total of 2.5 µs). A deep surface analysis of HER receptor revealed that the binding of specific anti-HER2 antibodies induced conformational changes both in the interfacial residues, which was expected, and in the ECDII, in particular at the "dimerization arm", which is critical in establishing protein-protein interface (PPI) interactions. Our results support and advance the knowledge on the already described trastuzumab effect on blocking HER2 dimerization through synergistic inhibition and/or steric hindrance. Furthermore, our approach offers a new strategy for fine-tuning target activity through allosteric ligands.

# Keywords

## Author summary

Increasing insight into the genetics and molecular biology of diseases has resulted in the identification of a high number of potential molecular targets for drug discovery and development. Human Epidermal Growth Factor Receptor 2 (HER2) is one of the most relevant Epidermal Growth Factor Receptor (EGFR) members, whose overexpression has been shown to play an important role in the development and progression of certain aggressive types of breast cancer. Thus, the development of novel approaches on anti-HER2 therapies is quiet relevant. Molecular modelling and simulation can be used to bring new perspectives, both in structure-based drug design and to provide atomistic information of the intermolecular coupling dynamics between inhibitors and receptors via interactive computographic software. Considering my interest in drug design and mechanisms of action using integrated in silico approaches, herein, I have used multiple methods to evaluate how HER2 coupling to its different partners could alter this functional mechanism. The results suggest that the antibodies fragments studied show different dynamic complexes with HER2 although both could contribute to downstream of the tumour cell receptor pathways. I do believe that future research breakthroughs with aid of chemo-bioinformatics will allow a more comprehensive perception on biomedicine.

## Introduction

The human epidermal growth factor receptor (EGFR) family comprises four members: human epidermal growth factor 1 (HER1), HER2, HER3 and HER4(1). They are overexpressed in a wide variety of malignant tumours, in particular in Non-Small Cell Lung Cancer (NSCLC) (40-80 %), breast cancer (14-91 %) and head and neck cancer (80-100 %) (2). In recent years, EGFR became an important biomarker and target of anti-tumour therapy(3). Within the EGFR family, HER2 is one of the most important members and when overexpressed, represents one of the most aggressive phenotypes (approximately 20-25 % of invasive breast cancers)(4), making HER2 a well-established therapeutic target. Trastuzumab (Herceptin®) was the first FDA-approved antibody against HER2 and remains the gold standard for the treatment of HER2-overexpressing

74    cancers(5). Clinical studies showed that the combination of trastuzumab with adjuvant

75    chemotherapy substantially improves disease-free survival (DFS) and overall survival(6).

76    From a structural point of view, HER2 is a 1255-amino-acid transmembrane glycoprotein with

77    185 kDa(7, 8) known for its function as a ligand-less co-receptor. It is noteworthy that HER2 is

78    the only member of the EGFR family that does not homodimerize under normal conditions, and

79    its ectodomain does not directly bind ligands(9). It is activated via heterodimerization with other

80    ligand-bound receptors, with HER3 the preferred heterodimerization partner(10) or

81    homodimerization when it is expressed at very high levels on the cell surface(11). HER2

82    dimerization activates cytoplasmic kinase, which leads to autophosphorylation of its intracellular

83    domains and thereby downstream signalling of cellular pathways. In normal cells, few HER2

84    molecules exist at the cell surface, and therefore few heterodimers are formed, allowing growth

85    signals to be relatively weak and controllable. However, if HER2 is overexpressed, the

86    dimerization level increases, resulting in enhanced responsiveness to growth factors and

87    malignant growth(8). Structurally, HER2 comprises an extracellular domain (~600 amino acid

88    residues, Fig 1A) with four extracellular domains (ECD) (ECDI-ECDIV), including two cysteine

89    rich domains (ECDII and ECDIV), a transmembrane domain, an intracellular tyrosine kinase

90    domain and a C-terminal tail with autophosphorylation sites(12).

91
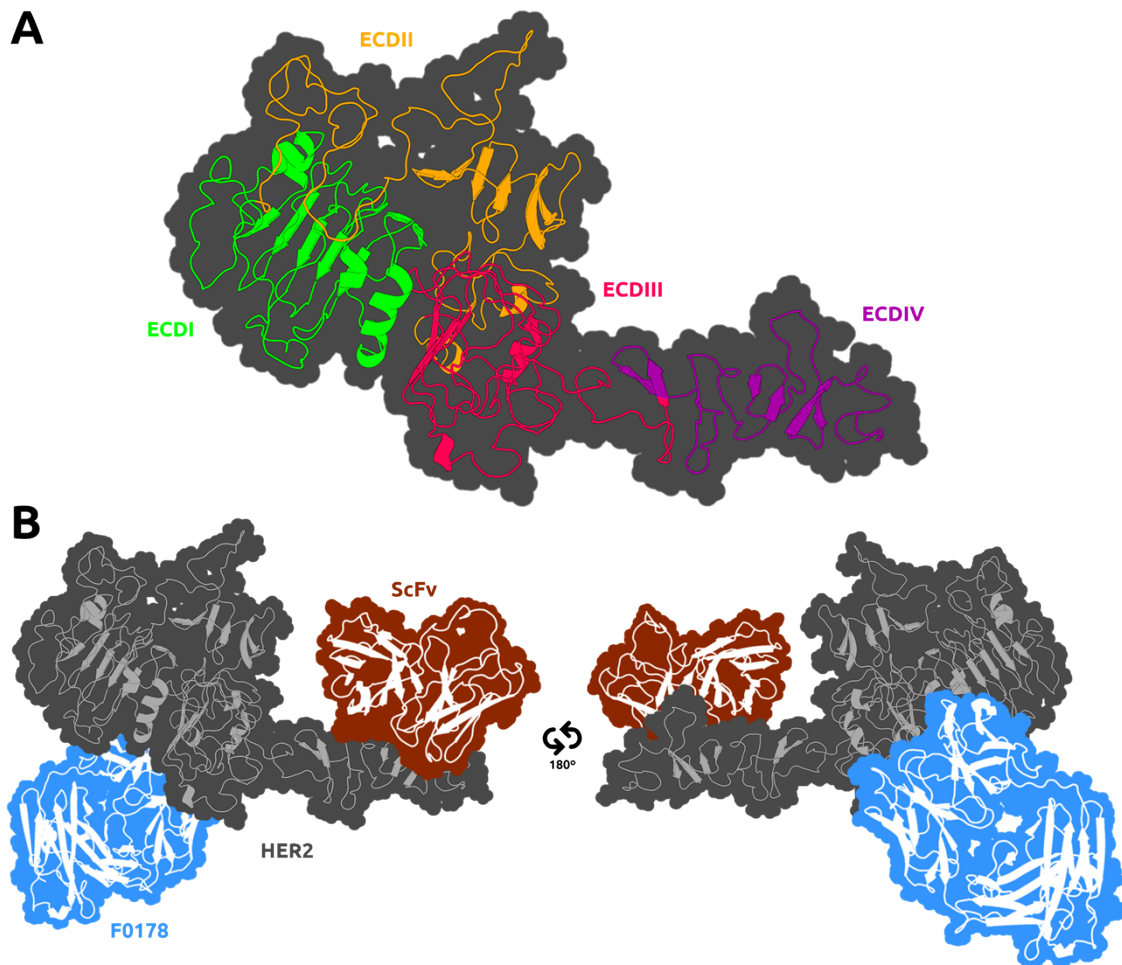
92
93 **Fig 1 -** Depiction of the HER2 receptor with its four different ECD regions mapped from PDBID:

94 1N8Z(13) (A). The regions where both partners are located at the complex crystal structure are

95 also shown (B). The F0178:HER2 structure was obtained from PDBID: 3WSQ(14) whereas the

96 ScFv:HER2 structure was obtained from PDBID: 1N8Z(13). The secondary structure of the

97 receptor and its partners is depicted as a cartoon, and the surface is shown as a contour.

98

99 The structures of the ligand-free HER2 ECDs (Fig 1A), ECDI and ECDIII, are very similar, with

100 each containing a "β-helix" with sides, formed by three parallel β-sheets due to a high number of

101 hydrophobic residues, mainly leucine. The spatial conformation of these two subdomains hinders

102 the effective access in part explaining lack of ligand binding by HER2. ECDII and ECDIV also

103 share a similar structure with small structural units held together by one or two disulphide bonds.

104 ECDIV, a near transmembrane domain, is crucial to the stabilisation of the protein-protein

105 interactions between HER2 and its dimerization partner(15). The structural conformation of ECDI

5

106    and ECDIII brings close together, allowing ECDII projection to participate in receptor

107    dimerization and initiation of signal transduction(16).   HER2 also includes a fixed, open

108    conformation overhand outside of the ECDII to contact, in particular, with the so-called

109    "dimerization arm" at $\Delta$265-288 region of its dimerization partner. This region is a short hairpin

110    loop(17, 18) with conserved residues and plays an important role in the dimerization process(16,

111    18, 19). Fu *et al.* have inferred the coupling mechanism of HER2 with the antigen-binding

112    antibody fragment (Fab) F0178 based on the three-dimensional (3D) static crystal structure of

113    F0178:HER2(14). The authors claimed that F0178 was able to bind to HER2 through a large and

114    highly complementary interface including both ECDI and ECDIII expressing a high

115    heterodimerization blocking activity(14).  However, the structural mechanism underlying HER2

116    dimerization is still unknown.

117    Here, we have employed a computational approach to deepen the current understanding of the

118    binding mechanism of HER2 with specific anti-HER2 antibody fragments. For this purpose, we

119    have selected two well-known anti-HER2-antibody fragments, namely F0178 and a single chain

120    variable fragment (scFv) from Trastuzumab. Both antibodies inhibit HER2 with distinct

121    mechanisms of action. The antibody fragment-HER2 complexes are depicted in Fig 1B. For the

122    study of the ScFv:HER2 system we have considered the crystal structure of HER2 with

123    trastuzumab Fab(13). The crystal structure showed that the trastuzumab Fab binds to the C-

124    terminal region of ECDIV of HER2, far away from the known dimerization region (ECDII). This

125    interaction blocks proteolytic cleavage in HER2 and indirectly affects dimerization with other

126    HERs(20). It has been hypothesised that binding of trastuzumab could induce conformational

127    changes providing a steric barrier avoiding kinase activation. This assembly could also promote

128    steric blocking of HER2 by proteolytic cleavage of its extracellular region and thereby altering

129    interaction between HER2 and other proteins(6).

130    *In silico* approaches are of crucial importance to clarify the structural and dynamical factors that

131    influence protein-protein interfaces (PPIs). They allow the prediction of how these interactions

132    occur, improving design and development of new therapeutic agents with better efficacies.

6

133   Aiming at a detailed picture of the overall coupling mechanism between the different antibody

134   fragments and the receptor as well as to understand the HER2 dimerization process, we employed

135   computational modelling and Molecular Dynamics (MD) simulations of HER2 receptor in the

136   apo form and upon complexation with F0178 and scFv from trastuzumab. We performed an

137   undocumented systematic analysis of structural, energetic and dynamic characteristics, ranging

138   from pairwise interactions formation to covariance analyses for these complexes. Moreover, we

139   identified the most important changes occurring at these PPIs and their influence on the

140   conformational rearrangement of the specific domains involved in dimerization, which is key to

141   the development of new HER2-specific therapies.

142

## Results

**1. Reaching equilibration in the MD simulations**

145   To attain equilibrium in MD simulations of large protein complexes is not a trivial task. We need

146   to evaluate the more conventional structural properties, like the root-mean-square deviation

147   (RMSD), but we will also need to monitor larger conformational transition of the quaternary

148   subunits that may appear at the slower timescales. The RMSD values were calculated using the

149   $C\alpha$ atoms of the receptor, with either using all ECD regions, or excluding ECDIV (Fig 2). The

150   inclusion of the ECDIV region in the calculations greatly increased the overall RMSD value due

151   to its high conformational flexibility as can be observed in unbound HER2 and in the

152   F0178:HER2 complex (Fig 2A-B). However, in ScFv:HER2, the contribution of ECDIV is

153   reduced, which was assigned to the strong interaction of the ScFv with this region (Fig 1B).

154   Overall, the ECDI-III region equilibrates relatively fast, while the ECDIV RMSD values indicate

155   larger movements occurring at larger timescales but of more difficult equilibration.

156
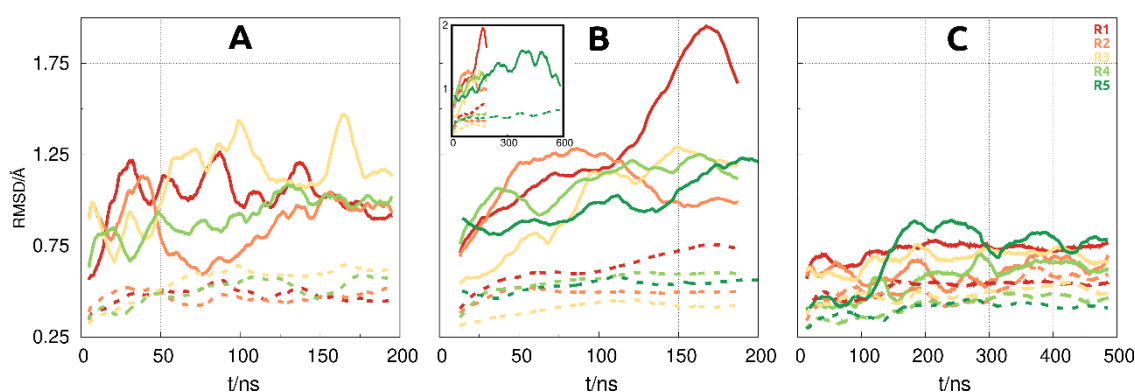
**Fig 2 -** RMSD plots of the five MD simulation runs (R1-R5) for HER2 (A), F0178:HER2 (B) and ScFv:HER2 (C). Full lines depict calculations made with all four ECD regions, whereas dotted lines do not include ECDIV. The small insert in Fig B shows the full length of the longer production run (R5).

To investigate the slower conformational transitions between HER2 and its partners, we have also calculated the centre-of-mass (COM) distance between them. This distance is almost invariant in F0178:HER2 (Fig S01), since F0178 is interacting directly with the ECDI-ECDIII region of HER2 (Fig 1B), inducing no significant domain movement. In the ScFv:HER2 system (Fig S02), we observed that the ScFv, upon binding, induced an opening/closing movement of the ECDIV towards the ECDI-III domains. As described in the methods section, this movement, which varies in the hundreds of nanoseconds time scale, was the reason why we have extended the MD simulations to improve its sampling.

Brought together these results showed that ScFv position varies between two extremes, depending on receptor proximity (Fig 3). The maximum COM distance leads to an opening of the cleft between ScFv and the HER2 ECDI-III domains (Fig 3A), an intermediate position, which is similar to the initial conformation (Fig 3B); and a minimum distance, which leads to cleft closure and a direct interaction of ScFv with ECDI-III (Fig 3C). Next, we selected the equilibrated regions of the MD simulations based on the RMSD and COM distances data. In the HER2 control and the F0178:HER2 complex simulations, these equilibrated rather fast (30-50 ns), whereas in the

8

179    ScFv:HER2 system, due to the large domain movements, some replicates took up to 150 ns to

180    reach equilibrium. In this system, all replicates seemed to converge to a different COM distance,

181    indicating that the use of multiple replicates was crucial to improve the sampling of this

182    movement.

183
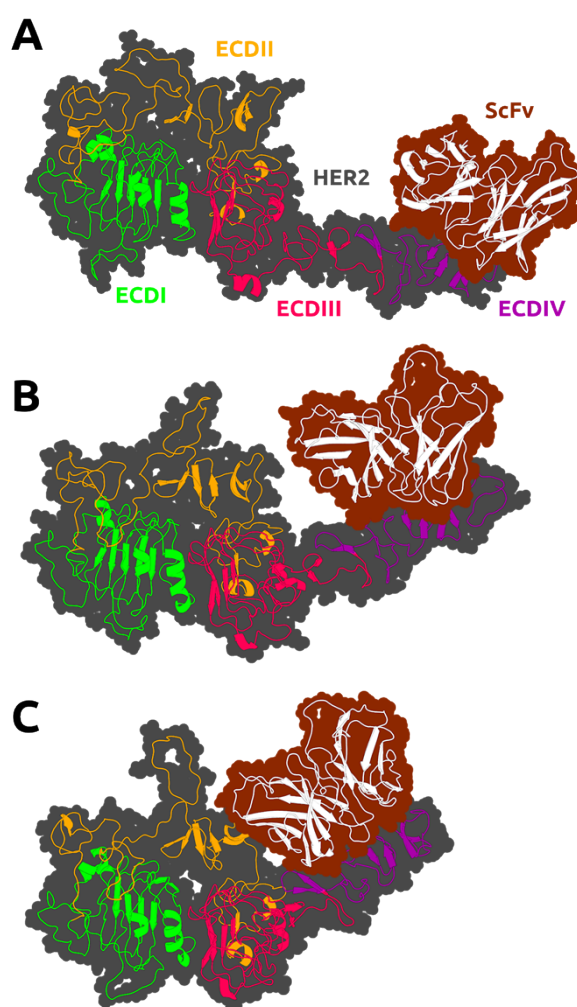


184

185    **Fig 3 -** Snapshots of the ScFv:HER2 system in open (A), intermediate (B) and closed (C)

186    conformations. The secondary structure of the receptor and its partners is depicted as a cartoon,

187    and the surface is shown as a contour.

188

189    **2. HER2 Conformational reorganisation**

190    After the equilibration evaluation of the MD simulations, we identified the equilibrium

191    conformational space and performed a deeper analysis of HER2 conformation reorganisation after

9

192     antibody binding. Briefly, it has been previously described that HER2-ECDII is the main domain

193     involved in the dimerization process(19, 21). Therefore, we have used an integrated approach of

194     multiple methods to evaluate how HER2 coupling to its different partners could alter this

195     functional mechanism. In addition, we have performed cross-correlation analysis for both
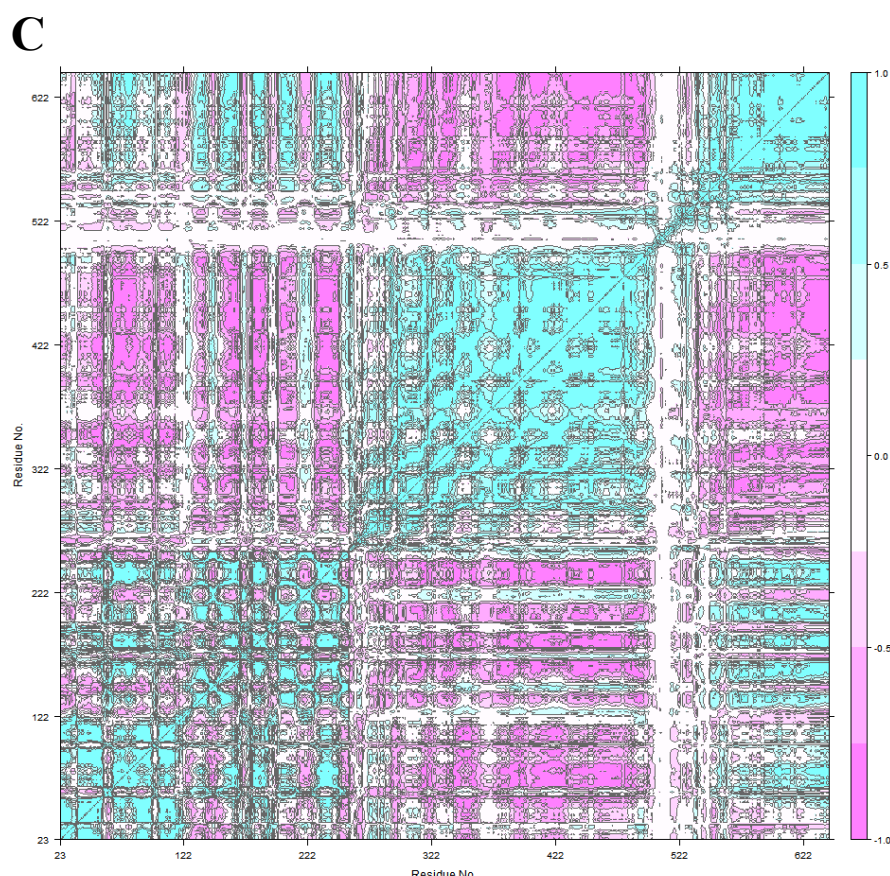
196     complexes (Fig 4).

197



198

199

11

**C**
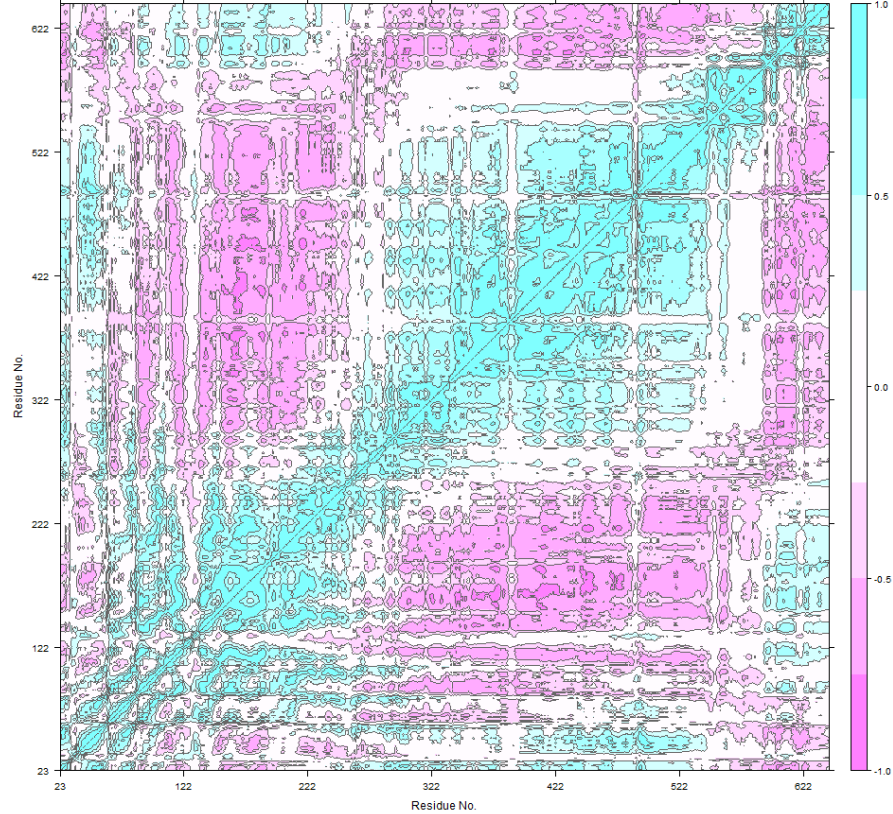


**Fig 4.** Cross-correlation Analysis. A: unbound HER2; B: HER2 at F0178:HER2 and C: HER2 at scFv:HER2.

For the unbound HER2 (Fig 4A) we found a mixed correlation spectrum (both positive and negative) for the 23-186 region (ECDI) and negative correlation values (excluding ECDIV) for the rest of the protein. Fu *et. al.* identified HER2 sub-region ASN498-ASP502 (ECDIII) as particularly altered due to F0178 binding even though it is positioned slightly apart(14). Cross-correlational analysis for F0178:HER2 shows strong and weak, respectively, negative correlation when comparing the 23-186 region (ECDI, especially sub-region 23-100) with ECDII (187-363) and ECDIII (364-530) regions. The negative correlation of ECDIII with ECDI (1-186 region) and ECDIV (531-626 region) underscores the importance of ECDIII on the F0178:HER2 interaction. Another very interesting aspect is the positive correlation between ECDIII and ECDII (187-363 region).

12

214    Concerning ScFv:HER2 (Fig 4C), a strong positive correlation between ECDIV tail (above 800)

215    and ECDI and ECDII revealed the importance of ECDIV on the ScFv:HER2 interaction. The

216    strong negative correlation of ECDII and ECDIII, as opposed to HER2 movement when is

217    unbounded (Fig 4A) suggested a conformational change with scFv binding at ECDIV. This

218    behaviour is in line with that shown in Fig 3 in which ScFv:HER2 complex varies its position

219    between two extremes.

220    To extract the relevant information from the conformational rearrangement observed during the

221    MD simulations, we performed a PCA analysis for the F0178:HER2 and ScFv:HER2 systems.

222    These calculations were performed using all replicas and considering only the equilibrated regions

223    of the MD simulations. Conformational clustering was also used to identify and further illustrate

224    the most populated conformers sampled in the MD simulations. For both systems, over 78-95 %

225    of the movement was described by the first 3 principal components, which allowed us to

226    positively identify the most relevant conformational transitions (Fig 5)

227



228

229

230

13
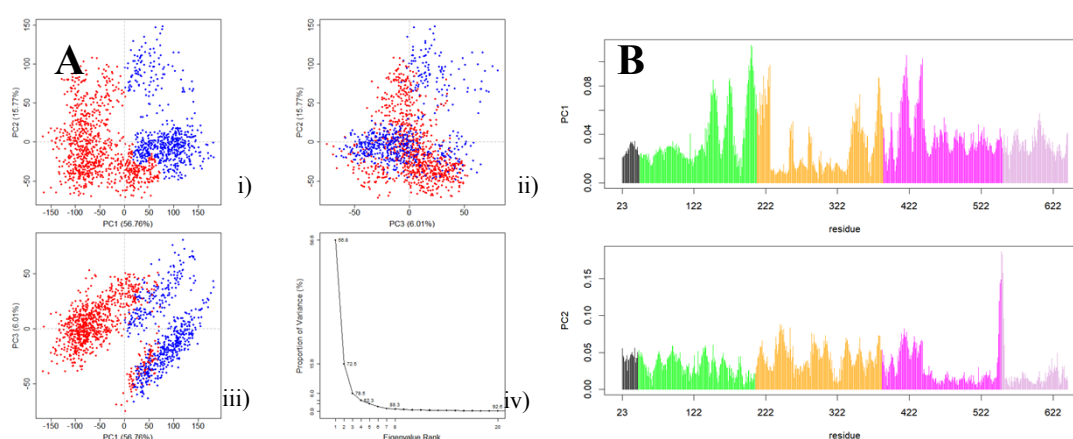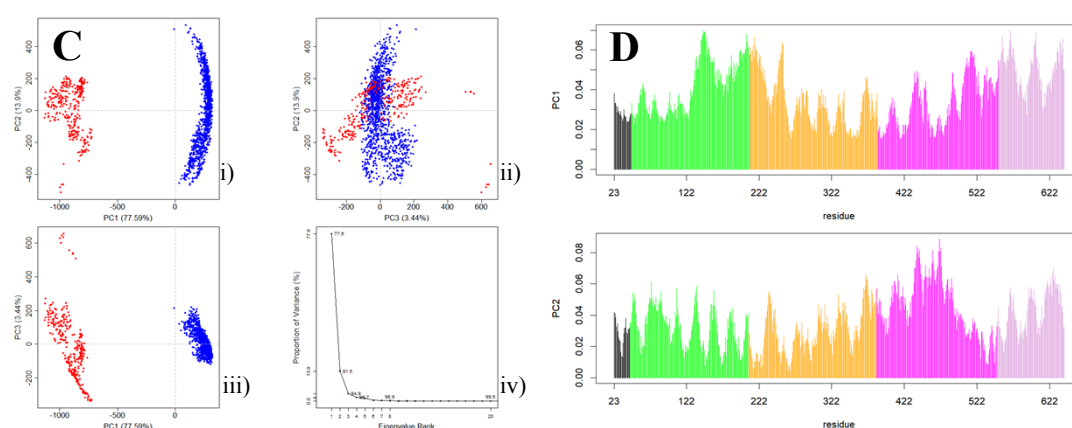
**Fig 5 -** PCA clustering results of the MD simulation trajectories of the complexes. i. Principal components (PC) 1 and 2. ii. PC1 and PC3. iii. PC2 and PC3. iv. Percentage of variance explained by the first 10 PC: (A) F0178:HER2, (C) scFv:HER2. Contribution of each residue to the first two principal components: (B) F0178:HER2, (D) scFv:HER2.

The PCA analysis for the F0178: HER2 MD simulations excluded the ECDIV tail, since its large movement (Fig 2B) would mask any other movement by the other domains. Assuming this approach allowed us to get insights about conformational details in regions otherwise inaccessible. Fig 5A shows slightly distinct conformational states mainly dependent on the ECDII at 187-363 region. Remarkably, the ECDII (187-363 region) residues had a significant contribution to the first two principal components (Fig 5B). We also calculated the conservation scores using the ConSurf web-server(22). Fig S03 shows a high conservation score for GLU352, VAL353 and ARG354 at F0178:HER2 complexes, which is in accordance with the higher contribution of ECDII region for PC1.

PCA analyses suggested two main conformational states for the scFv: HER2 MD simulations (Fig 5C). In particular, the contribution of the ECDII residues to PC1 and PC2 (Fig 5D) suggested that the dimerization region was indeed affected by binding. Fig S03 also supports this behaviour with a high rate of the conservation pattern at ECDII, namely at dimerization arm (GLU265-ARG288). Paradoxically, the role of the ECDII-HER2 residues in the interface-interaction network of both complex systems were not considered in the literature (13, 14). To further aid interpretation, we

14

252  produced a trajectory that interpolates between the most dissimilar structures in the distribution

253  along PC1 for both systems (Fig S04). The movement of the "dimerization arm" at ECDII is rather

254  evident, as well as at ECDI. The role of this region for HER2 coupling mechanism has not been

255  previously described although the blocking ability of HER2 heterodimerization is well

256  recognised(14). We propose that the HER2 ECDII conformational change is coupled with this

257  allosteric regulation of F0178:HER2.

258  Concerning the scFv:HER2 complex system, we observed an approximation between ECDII and

259  scFv compared to the initial 3D crystallographic structure (PDBID 1NZ8(13)) (Fig S04). This is

260  an important biological feature since it is well-known that the 'heart-shaped' domain II open

261  conformation is essential for dimerization[17].

262

263  **3. The impact of the ScFv and F0178 on the conformational space of HER2**

264  Based on the analysis of the conformational changes that occurred during the MD simulations of

265  the systems, the ScFv and F0178 antibody fragments seem to induce conformational changes in

266  HER2, which might alter its dimerization propensity. The most impactful changes are those that

267  alter the receptor surface, in particular, in the regions known to be involved in dimerization. These

268  surface changes are inevitably coupled with variations in the solvent exposure of certain residues.

269  Therefore, we calculated the solvent accessible surface area (SASA) of each HER2 residue in all

270  systems to calculate significant differences between complexes and the unbound receptor. The

271  residues with significant differences in SASA (difference > 5 % and standard error is smaller than

272  the difference value) due to the presence of a partner are associated with a direct interaction and

273  should correlate well with the interfacial region (Fig 6). As expected, the residues with the largest

274  shift in SASA in F0178:HER2 are located in ECDI, ECDII and ECDIII (Fig 1B and Fig 6A),

275  whereas in ScFv:HER2 these residues are mainly located in ECDII and ECDIV (Fig 1B and Fig

276  6B). In the case of ScFV:HER2, the direct interaction of the ligand with ECDII is due to the cleft

277  closure (Fig 3C), observed in several simulations (Fig S02). The large error bars in F0178:HER2

278    were attributed to insufficient sampling and/or an increase variability in the interfacial region of

279    the complex.

280
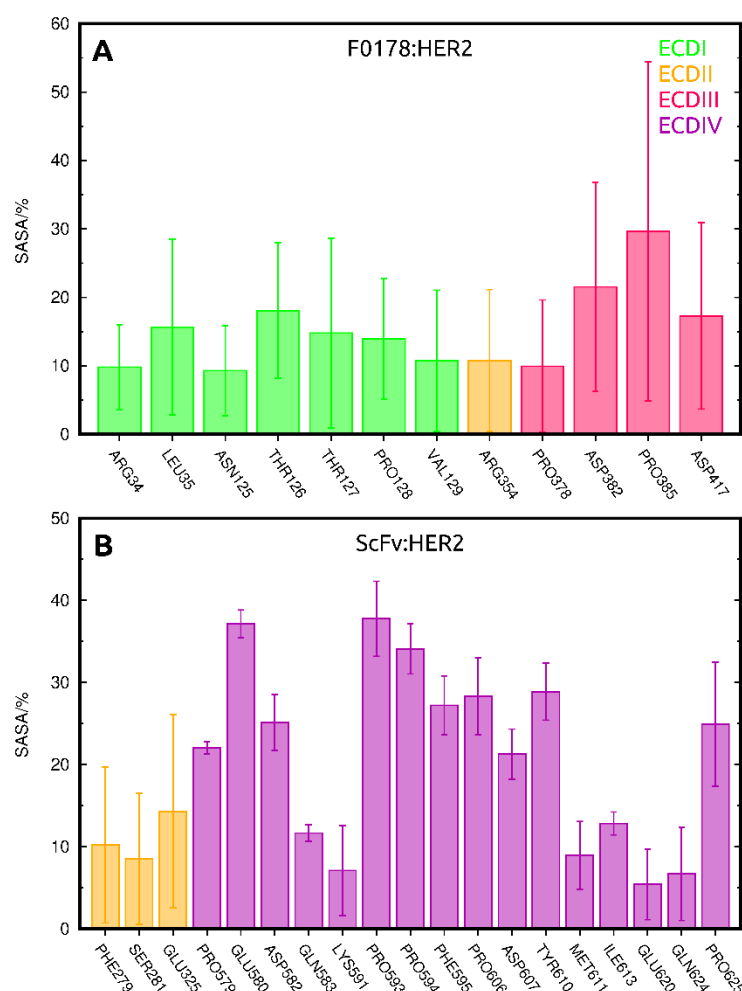


281

282    **Fig 6 -** Residues with significant differences in SASA due to the presence of a partner, coloured

283    by their respective ECD in F0178:HER2 (A) and ScFv:HER2 (B).

284

285    We also identified several residues whose SASA changed between unbound HER2 and its

286    complexes through an indirect, or allosteric, effect of the antibody fragment. This was attained by

287    calculating the SASA differences between F0178:HER2/ScFv:HER2 and HER2 trajectories,

288    ignoring the physical presence of the ligands. The residues identified with this procedure (Fig 7

289    and Fig S05) were different in both complexes, but located predominantly in ECDII, including

290    the so-called "dimerization arm" region (TYR274, GLU280 and MET282 in F0178:HER2; and

291    CYS268 in ScFv:HER2).

16

292    Our results indicate that F0178 and scFv form fairly distinct dynamic complexes with HER2. The

293    binding of F0178 seems to provoke a marked allosteric effect on the "dimerization arm" residues

294    at ECDII, which is the dimerization region of HER2(16). A similar allosteric effect could be

295    observed for scFv, which was complemented with a direct interaction with the ECDII region.

296    This interaction map obtained from our simulations may help explain the strong effect of scFv on

297    HER2 homodimerization(6), opening new therapeutic avenues for targeting HER2 inhibition.

298

299
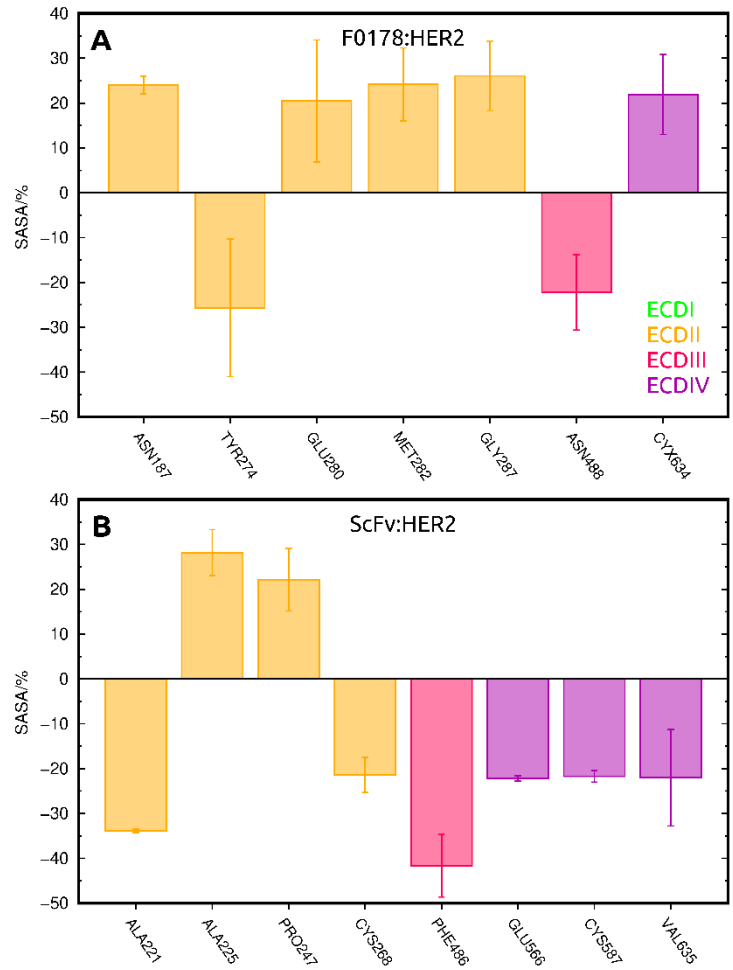


301    **Fig 7** - Residues with the largest difference in SASA due to allosteric effects, colored by their

302    respective ECD in F0178:HER2 (top) and ScFv:HER2 (bottom).

303

## Discussion and Conclusions

305    HER2 plays a key role in the HER family due to its specificity to interact with other HER receptors

306    through a complex signalling network to regulate cell growth, differentiation and survival. HER2

307    overexpression occurs in approximately 15–30 % of breast cancers and 10–30 % of

308    gastric/gastroesophageal cancers and serves as a prognostic and predictive biomarker. The easy

309    accessibility to the four HER2 extracellular domains (ECD) turns this receptor into a highly

310    relevant target for the selective delivery of anti-tumour drugs as well as imaging agents. ECDI

311    and ECDIII can form a binding site for potential HER2 ligands, whereas the cysteine-rich ECDII

312    and ECDIV are known to be involved in the homodimerization and heterodimerization processes.

313    Furthermore, ECDII, in particular the "dimerization arm" region (Δ265-288), is believed to be the

314    main contributor to dimerization. Although crystal structures are known, the dynamics of the

315    binding of different antibodies to this receptor has remained unclear. In this work, we have applied

316    an integrated framework of molecular modelling tools and techniques as well as MD simulations

317    analyses to address important structural factors that provide atomistic information of the

318    intermolecular coupling dynamics between HER2 and two different HER2 inhibitors, namely

319    F0178 Fab and scFv from trastuzumab. Both are well-studied anti-HER2 antibody fragments that

320    act at different ECD domains of HER2. F0178 interacts with both ECDI and ECDIII domains,

321    while scFv binds at ECDIV domain. Concerning the ScFv:HER2 complex, the highest

322    contribution to the PPI stability occurs in the terminal part of ECDIV.

323    In addition to the detailed residue and pairwise interaction analyses of HER2-antibody complexes,

324    an important contribution of our work is on the conformational changes that occur upon binding.

325    Our simulations suggested that F0178 and scFv form distinct dynamic complexes with HER2.

326    The binding of F0178 seems to induce a marked allosteric effect on the "dimerization arm",

327    whereas scFv binding leads to major conformational changes on HER2 structure (ECDI

328    approaches ECDII and ECDIII). We observed a drastic movement of ECDII towards scFv,

329    establishing a new possible interface capable of meaningful and important interactions with the

330    antibody. This newly discovered conformational organisation of HER2 upon scFv coupling could

331    indeed explain its role on dimerization as suggested by Hudis *et al.*(6) and herein we give a

18

332     pioneer view of this crucial dynamical process. The integrated analysis framework applied here

333     has unveiled the role of F0178 and scFv from trastuzumab in HER2 dimerization. Our results

334     provide a new perspective for future approaches to the selection of effective potential drugs, since

335     the conformational changes triggered by the ligands can have positive effects downstream of the

336     tumour cell receptor pathways.

337

## Experimental section

### 1. Model Construction

340     The 3D structure of HER2 extracellular domain was constructed by homology modelling using

341     the MODELLER package(23), and the protein template NP_004439.2 (PDBID: 3N85A)(24) with

342     the sequence 301598662. The loop regions LEU99:SER133; THR245:CYS264 and

343     ASP121:GLY136 were further refined with ModLoop(25, 26). The complexes of HER2 with

344     antibody fragments were retrieved from the correspondent PDB files, specifically F0178:HER2

345     from PDBID: 3WSQ(14) and ScFv:HER2 from trastuzumab with PDBID: 1N8Z(13). The

346     protonation states of residues in the physiological pH range were determined with the PROPKA

347     methodology (27-29), which computes $pK_a$ values of ionisable residues by accounting for the

348     effect of the protein environment. PKA2PQR(27) was applied to convert PDB files in PQR files

349     by AMBER ff99 force field(30). All histidine residues were considered deprotonated at

350     physiological pH conditions save for HIS 349, which was assigned as positively charged.

### 2. Molecular Dynamics Simulations

352     MD simulations of HER2, F0178:HER2 and ScFv:HER2 were performed using GROMACS

353     2018.3(31-33) and the Amber ff99SB-ILDN force field (34). Each complex was solvated by a

354     TIP3P water box with 12 Å buffering distance to the boundary, which was verified throughout

355     the simulations. The simulation systems were kept neutral by adding the necessary counterions.

356     Simulations were performed in the *NPT* (isothermal-isobaric) ensemble. Temperature coupling

357     was performed using the v-rescale thermostat(35) at 300 K, with a coupling constant of 0.1 ps,

19

358     while an isotropic Parrinello-Rahman barostat(36, 37)  was used to keep the pressure constant at

359     1 bar using a coupling constant of 2.0 ps and a compressibility of $4.5 \times 10^{-5}$ $bar^{-1}$. Long range

360     electrostatic interactions were computed using the particle mesh Ewald (PME) method (38, 39)

361     with a Fourier grid spacing of 0.16 nm and a cut-off of 1.0 nm for direct contributions. Lennard-

362     Jones interactions were computed using a nonbonded neighbour pair list with a cut-off of 1.0 nm,

363     enabling the use of the Verlet scheme (40). Solute bonds were constrained using the Parallel

364     LINear Constraint Solver, P-LINCS (41), and solvent molecules were constrained using SETTLE

365     algorithm(42). The initial system energy was first minimised for 50.000 steps, followed by two

366     equilibrations of 250 ps.  The HER2 system was then initialised for 10 ps using position restraints

367     on all protein heavy atoms with a force constant of 1000 kJ $nm^{-2}$ $mol^{-1}$, while the remaining

368     systems were initialised for 25 ps without restraints.

369     Initially, four production runs were performed for each system, each 200 ns-long. In addition, for

370     the F0178:HER2 and ScFv:HER2 complexes, a longer production run was also performed (~500

371     ns-long). However, when analysing the centre-of-mass (COM) distance between the partners in

372     the HER2:ScFv system (Fig S02), we decided to extend the length of all production runs to 500

373     ns, in order to obtain better sampling of this motion.

374     The equilibrated regions for each system (Table S03, in the Supporting Information) was

375     replicate/system dependent, and the decision was based on the RMSD and COM distance

376     variations. The RMSD calculations were performed using the Cα atoms either in all extracellular

377     domains, or in the ECDI-ECDIII region, as shown in Fig 2. The COM distance plots (Fig 3 and

378     Figs S01-02) were obtained by calculating the centre-of-mass of the HER2 and either ScFv or

379     F0178, using all atoms in the receptor and each of its partners.

380

381         **3. Solvent-accessible surface area analysis**

382     We calculated the solvent-accessible surface area (SASA) of each individual residue in HER2,

383     ScFv:HER2 and F0178:HER2 using the equilibrated regions of each replicate. In the unbound

384     HER2, for instance, this resulted in four average values of SASA per residue, which were then

385  used to calculate the mean and standard error of the mean (SEM) using a jackknife technique. We

386  then converted these absolute SASA values into percentages by calculating the maximum value

387  a fully exposed residue of each type would have. With these percentages, we excluded all residues

388  whose solvent-access was below 20 %, as they were considered mainly internalised.

389

### 3.1 SASA analysis of interfacial residues

391  The identification of residues in the interface between HER2 and each of its partners was

392  performed by calculating the SASA of each residue of the receptor in the complex, with and

393  without its partner (e.g., the residue coloured in yellow depicted in Fig S06). When comparing

394  the SASA values of each complex in the presence and absence of the antibody, we were able to

395  exclude all residues whose SASA was not altered, *i.e.* residues which are not interacting directly

396  with the partner. Finally, we only reported residues whose difference in SASA was above 5 %

397  and whose SEM was lower than this difference.

398

### 3.2 SASA analysis of allosterically modulated residues

400  To identify allosterically modulated residues, we compared the SASA of all residues in each

401  complex with the same residue in unbound HER2. This SASA difference identified which

402  residues became more exposed or occluded by an indirect effect of the antibody (*e.g.* the residue

403  coloured in green depicted in Fig S06). As in the previous protocol, residues whose difference in

404  SASA between systems was below 5 % and whose SEM was higher than the difference itself

405  were excluded.

406

### 4. Principal Component Analysis

408  Principal Component Analysis (PCA)(43) is a useful statistical technique able to reveal the more

409  important patterns in a dataset. The most relevant information from a dataset are extracted as

410  eigenvalues or principal components (PCs). Internal correlation motions between residues of

21

411 interface complexes were investigated by performing a cross-correlations analysis on the MD

412 trajectory given a covariance matrix in Cartesian coordinate space by Eq. (1).

413 $$c_{ij} = \; < (x_i - <x_i>)(x_j - <x_j>) >$$ Eq. (1)

414 Where $x_i$ and $x_j$ are coordinates of the $i$th and $j$th atoms in the system and $< \dots >$ denotes trajectory

415 average[44]. The matrix $c_{ij}$ symmetric can be diagnosed by using orthogonal coordinate

416 transformation matrix T, transforming the matrix $c_{ij}$ into a diagonal matrix $\Lambda$ of eigenvalues $\lambda_i$ by

417 Eq. (2):

418 $$\Lambda = T^T c_{ij} T$$ Eq. (2)

419 Where each eigenvector of the matrix gives the direction, along which a concerted motion arises,

420 and the eigenvalue gives the magnitude of fluctuations along this direction. The PCA calculations

421 were performed using bio3d[45].

422

423 **5. Evolutionary Conservation of Interfacial Residues**

424 Sequence conservation was retrieved for all interfacial residues in HER2 using Consurf [46],

425 which is based on the Rate4Site algorithm[47]. MAFFT[48, 49] was used for multiple sequence

426 alignment (MSA), using the BLAST[50] on the UNIREF90 database[51]. Each MSA comprised

427 at most 150 sequences, with homology values ranging from 35 % to 95 %.

428

448

# References

1.      Wieduwilt MJ, Moasser MM. The epidermal growth factor receptor family: Biology driving targeted therapeutics. Cellular and molecular life sciences : CMLS. 2008;65(10):1566-84.

2.      Herbst RS. Review of epidermal growth factor receptor biology. International Journal of Radiation Oncology*Biology*Physics. 2004;59(2, Supplement):S21-S6.

3.      Zhou X, Li J, Wang Z, Chen Z, Qiu J, Zhang Y, et al. Cellular Immunotherapy for Carcinoma Using Genetically Modified EGFR-Specific T Lymphocytes. Neoplasia (New York, NY). 2013;15(5):544-53.

4.      Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science. 1987;235(4785):177-82.

5.      Vu T, Claret FX. Trastuzumab: updated mechanisms of action and resistance in breast cancer. Frontiers in oncology. 2012;2:62-.

6.      Hudis CA. Trastuzumab — Mechanism of Action and Use in Clinical Practice. New England Journal of Medicine. 2007;357(1):39-51.

7.      Coussens L, Yang-Feng T, Liao Y, Chen E, Gray A, McGrath J, et al. Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with neu oncogene. Science. 1985;230(4730):1132-9.

8.      Rubin I, Yarden Y. The Basic Biology of HER2. Annals of Oncology. 2001;12(suppl_1):S3-S8.

9.      Lv Q, Meng Z, Yu Y, Jiang F, Guan D, Liang C, et al. Molecular Mechanisms and Translational Therapies for Human Epidermal Receptor 2 Positive Breast Cancer. International journal of molecular sciences. 2016;17(12):2095.

10.      Olayioye MA, Neve RM, Lane HA, Hynes NE. The ErbB signaling network: Receptor heterodimerization in development and cancer. EMBO Journal. 2000;19(13):3159-67.

11.      Gutierrez C, Schiff R. HER2: biology, detection, and clinical implications. Archives of pathology & laboratory medicine. 2011;135(1):55-62.

12.      Moasser MM. The oncogene HER2; Its signaling and transforming functions and its role in human cancer pathogenesis. Oncogene. 2007;26(45):6469-87.

13.      Cho H-S, Mason K, Ramyar KX, Stanley AM, Gabelli SB, Denney DW, et al. Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. Nature. 2003;421(6924):756-60.

14.      Fu W, Wang Y, Zhang Y, Xiong L, Takeda H, Ding L, et al. Insights into HER2 signaling from step-by-step optimization of anti-HER2 antibodies. mAbs. 2014;6(4):978-90.

15.      Landgraf R. HER2 therapy. HER2 (ERBB2): functional diversity from structurally conserved building blocks. Breast cancer research : BCR. 2007;9(1):202-.

16.      Hu S, Sun Y, Meng Y, Wang X, Yang W, Fu W, et al. Molecular architecture of the ErbB2 extracellular domain homodimer. Oncotarget. 2015;6(3):1695-706.

17.      Bessman NJ, Bagchi A, Ferguson KM, Lemmon MA. Complex relationship between ligand binding and dimerization in the epidermal growth factor receptor. Cell reports. 2014;9(4):1306-17.

18.      Arkhipov A, Shan Y, Kim ET, Dror RO, Shaw DE. Her2 activation mechanism reflects evolutionary preservation of asymmetric ectodomain dimers in the human EGFR family. eLife. 2013;2:e00708.

19.      Tai W, Mahato R, Cheng K. The role of HER2 in cancer therapy and targeted drug delivery. Journal of controlled release : official journal of the Controlled Release Society. 2010;146(3):264-75.

20.      Banappagari S, Ronald S, Satyanarayanajois SD. A conformationally constrained peptidomimetic binds to the extracellular region of HER2 protein. Journal of biomolecular structure & dynamics. 2010;28(3):289-308.

499  21.     Pietras RJ, Arboleda J, Reese DM, Wongvipat N, Pegram MD, Ramos L, et al. HER-2
500  tyrosine kinase pathway targets estrogen receptor and promotes hormone-independent growth
501  in human breast cancer cells. Oncogene. 1995;10(12):2435-46.

502  22.     Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, et al. ConSurf: identification
503  of functional regions in proteins by surface-mapping of phylogenetic information.
504  Bioinformatics. 2003;19(1):163-4.

505  23.     Pieper U, Eswar N, Braberg H, Madhusudhan MS, Davis FP, Stuart AC, et al. MODBASE,
506  a database of annotated comparative protein structure models, and associated resources.
507  Nucleic Acids Research. 2004;32(Database issue):D217-D22.

508  24.     Fisher RD, Ultsch M, Lingel A, Schaefer G, Shao L, Birtalan S, et al. Structure of the
509  Complex between HER2 and an Antibody Paratope Formed by Side Chains from Tryptophan and
510  Serine. Journal of Molecular Biology. 2010;402(1):217-29.

511  25.     Fiser A, Do RK, Sali A. Modeling of loops in protein structures. Protein Science : A
512  Publication of the Protein Society. 2000;9(9):1753-73.

513  26.     Fiser A, Sali A. ModLoop: automated modeling of loops in protein structures.
514  Bioinformatics. 2003;19(18):2500-1.

515  27.     Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, et al. PDB2PQR: expanding
516  and upgrading automated preparation of biomolecular structures for molecular simulations.
517  Nucleic Acids Research. 2007;35(Web Server issue):W522-W5.

518  28.     Søndergaard CR, Olsson MHM, Rostkowski M, Jensen JH. Improved Treatment of
519  Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. Journal
520  of Chemical Theory and Computation. 2011;7(7):2284-95.

521  29.     Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent
522  Treatment of Internal and Surface Residues in Empirical pKa Predictions. Journal of Chemical
523  Theory and Computation. 2011;7(2):525-37.

524  30.     Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP)
525  model perform in calculating conformational energies of organic and biological molecules?
526  Journal of Computational Chemistry. 2000;21(12):1049-74.

527  31.     Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel
528  molecular dynamics implementation. Computer Physics Communications. 1995;91(1):43-56.

529  32.     Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast,
530  flexible, and free. J Comput Chem. 2005;26(16):1701-18.

531  33.     Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High
532  performance molecular simulations through multi-level parallelism from laptops to
533  supercomputers. SoftwareX. 2015;1-2:19-25.

534  34.     Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved
535  side-chain torsion potentials for the Amber ff99SB protein force field. Proteins. 2010;78(8):1950-
536  8.

537  35.     Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem
538  Phys. 2007;126(1):014101.

539  36.     Nosé S, Klein ML. Constant pressure molecular dynamics for molecular systems.
540  Molecular Physics. 1983;50(5):1055-76.

541  37.     Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular
542  dynamics method. Journal of Applied Physics. 1981;52(12):7182-90.

543  38.     Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums
544  in large systems. The Journal of Chemical Physics. 1993;98(12):10089-92.

545  39.     Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle
546  mesh Ewald method. The Journal of Chemical Physics. 1995;103(19):8577-93.

547  40.     Páll S, Hess B. A flexible algorithm for calculating pair interactions on SIMD architectures.
548  Computer Physics Communications. 2013;184(12):2641-50.

549  41.     Hess B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. Journal of
550  Chemical Theory and Computation. 2008;4(1):116-22.

42.	Miyamoto S, Kollman PA. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. Journal of Computational Chemistry. 1992;13(8):952-62.

43.	Abdi H, Williams LJ. Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics. 2010;2(4):433-59.

44.	Yang M, Zhang X, Han K. Molecular dynamics simulation of SRP GTPases: Towards an understanding of the complex formation from equilibrium fluctuations. Proteins: Structure, Function, and Bioinformatics. 2010;78(10):2222-37.

45.	Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics. 2006;22(21):2695-6.

46.	Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 2016;44(W1):W344-50.

47.	Pupko T, Bell RE, Mayrose I, Glaser F, Ben-Tal N. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. Bioinformatics. 2002;18(suppl_1):S71-S7.

48.	Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research. 2002;30(14):3059-66.

49.	Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:113.

50.	Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10.

51.	Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, UniProt C. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics. 2015;31(6):926-32.

## Supporting information captions

**Fig S01 - Time evolution of the COM distance between HER2 and F0178.** The smaller plot shows the full length of the larger production run for this system. Data was smoothed using a floating average window of 10 ns.

**Fig S02 - Time evolution of the COM distance between HER2 and ScFv.** Data was smoothed using a floating average window of 10 ns.

**Fig S03 - Normalised conservation scores of interfacial residues from ECDII-HER2 when coupled to F0178 (•) and scFv (+) antibody fragments.** The light blue bars highlight the so-called "dimerization arm" residues. The light blue and dark blue dashed lines represent the average conservation score for dimerization arm and non-dimerization arm residues, respectively.

587 **Fig S04. Interpolated structures along PC1: (A) F0178:HER2, (B) scFv:HER2.** Colour scale

588 from blue to red depict low to high atomic displacements. The dashed box highlights the

589 "dimerization arm" region at ECDII.

590 **Fig S05 - Location of the residues that showed a significant shift in SASA due to allosteric**

591 **effects.** The regions coloured in red show the residues in HER2 whose SASA was affected by the

592 presence of ScFv, while the regions in blue show the residues in HER2 whose SASA was affected

593 by F0178. Also shown are the initial positions of both partners (semi-transparent). The image on

594 the right is the mirror-image of the one on the left. The secondary structure of the receptor and its

595 partners is depicted as a cartoon, and the surface is also shown as a contour.

596 **Fig S06 - Simplified depiction of the SASA analysis.** The dotted lines are a simplified depiction

597 of the region used in the SASA calculation. Two specific residues coloured in green and yellow

598 are also included. The residue in yellow is located in the interface region between HER2 and ScFv

599 and should have a noticeable difference in SASA if ScFv is included or excluded from the

600 calculation in the complex trajectory. The residue in green is located in a region where the

601 presence of ScFv could only be felt indirectly when performing differences between the bound

602 and unbound trajectories

603 **Table S01 - MD equilibrated regions for each system and replicate in nanoseconds.**

604