Structure-based prediction of Ras-effector binding affinities and design of 'branchegetic' interface mutations

Philipp Junk [2,3*], Christina Kiel [1,2,3]

Affiliations

[1] Department of Molecular Medicine, University of Pavia, 27100 Pavia, Italy

[2] Systems Biology Ireland, School of Medicine, University College Dublin, Dublin 4, Ireland.

[3] UCD Charles Institute of Dermatology, School of Medicine, University College Dublin, Dublin 4, Ireland

*Correspondence: Philipp Junk (philipp.junk@ucdconnect.ie)

Summary

Ras is a central cellular hub protein controlling multiple cell fates. How Ras interacts with a variety of potential effector proteins is relatively unexplored, with only some key effectors characterized in great detail. Here, we have used homology modelling based on X-ray and AlphaFold templates to build structural models for 54 Ras-effectors complexes. These models were used to estimate binding affinities using a supervised learning regressor. Furthermore, we systematically introduced Ras 'branch-pruning' (or branchegetic) mutations to identify 200 interface mutations that affect the binding energy with at least one of the model structures. The impacts of these branchegetic mutants were integrated into a mathematical model to assess the potential for rewiring interactions at the Ras hub on a systems level. These findings have provided a quantitative understanding of Ras-effector interfaces and their impact on systems properties of a key cellular hub.

Keywords

AlphaFold, branch pruning, edgetics, enedgetics, effectors, FoldX, homology modelling, hub, networks, Ras

Introduction

Ras is a key cellular signaling hub and oncogene (*1*). The first correct Ras structure was famously described in 1989 (*2*) (*3*) and consists of the G domain super-fold (six β-strands and five α-helices; (*1*)). It's main structural and functional characteristic is the nucleotide binding site (the typical α,β-fold of nucleotide binding proteins), which can bind GTP and hydrolyze it to GDP. Then, GDP gets released and a new nucleotide (favorably GTP since it is higher abundant in cells) gets bound. Depending on which nucleotide is bound, the functional state of RAS is different. GDP bound Ras assumes a so-called inactive conformation, whereas upon the binding of GTP, the conformation of Ras is called active. The difference between the active and the inactive conformation is that, in the active conformation, two loop regions called switch 1 and switch 2 are interacting with the GTP and thereby tightly bound to it (*1*) (Figure 1A). This rearranges the interface of Ras in a specific way so that effector proteins with a specific structural motive can interact with Ras. The structural motive required for the interaction with Ras has a ubiquitin domain super-fold (*4*). There exist three families of these domains, called Ras binding domain (RBD), Ras association domain (RA domain) and PI3K-Ras binding domain (PI3K RBD) with only minor sequence and structural differences between them. In the following, for the sake of simplicity, all of these will be called RBDs.

Ras·GTP can interact and activate many downstream effectors, some of which are better studied than others (*5*). Well-studied effectors include PI3-kinases, RalGDS, and Raf kinases, which control important cellular processes such as survival, polarization, adhesion, migration, and proliferation. In our previous work, we characterized a set of 56 RBD-containing proteins as potential Ras effectors, which converge into 12 classes that are linked to different downstream cellular processes and phenotypes (*6*) (*7*) (*8*) (*9*).

Ras-effector interactions are interesting from a systems biology and network point of view. Effectors use a mutually exclusive binding site on Ras, hence competition for binding can occur under certain conditions (*10*). Also, the experimental and predicted affinities between RBDs and Ras·GTP vary, ranging from nano- to micromolar K_d values (*8*). Previously we analyzed how the amount of effectors in complex with Ras varies in different cell types (*6*) (*7*). We find that only 9 effectors are predicted to bind in significant amount to Ras·GTP using the RBD alone (*7*). However, 31 effectors are predicted to form significant complexes with Ras if they are additionally recruited to the plasma membrane via other domains present in effectors (piggyback mechanism, (*11*)). Hence, even weak binding affinities on the level of RBD-RAS binding can turn into significant complex formation if effectors are recruited (e.g. in a context

65 specific manner). Thus, for systems analyses and computational models, we need good

66 estimations for all binding affinities between RBDs and Ras.

67 Structural analysis can lead to a deeper understanding of protein-protein interaction

68 (PPI). Previous we used homology modelling, based on experimentally determined complex

69 structures, to predict binding between effectors and Ras·GTP (*12*) (*13*). While this has provided

70 insights into binding propensities, a limitation of this work was that no quantitative binding

71 affinities were predicted, but qualitative binding by classifying effectors into categories of

72 'binding', 'non-binding', and 'twilight'. It is now timely to revisit homology modelling of Ras-

73 RBD complexes, as new template structures are available by (i) X-ray crystallography (8 of 56

74 effectors are crystallized in complex with RAS) and (ii) AlphaFold (*14*) (*15*).

75 Ras is frequently mutated in different cancers, where aberrant Ras signaling plays a role

76 in cancer initiation, progression, and metastasis via alterations in metabolism, proliferation, and

77 survival (*16*). As Ras cannot be directly targeted (or only certain Ras mutations) (*17*), much

78 hope rests on network-centric approaches (*18*) that involve Ras-effector interactions or

79 downstream pathways (*9*). Therefore, tinkering with Ras-effector binding is an attractive

80 alternative strategy for finding suitable targets for therapeutic interventions. Previously we

81 developed a 'branch-pruning' (or "branchegetic", in analogy to "edgetics" (*19*) and

82 "enedgetics" (*20*)) strategy for Ras effector interactions, where mutations are introduced into

83 Ras that differentially impact binding to effectors (*21*). For example, introducing a mutation

84 can result in a steric clash in the interface formed with one effector, but not with another; hence

85 the interaction with one effector is broken while intact for another.

86 In this work, we first generate homology models of all Ras-effector (RBD) interactions

87 and predict affinities of the RBDs in complex with Ras·GTP. We then use the generated model

88 structures to employ a systematic branchegetic strategy that explores the impact of Ras interface

89 mutants on binding to all effectors. Altogether, our results contribute to a quantitative and

90 systems-level understanding of Ras-effector interactions and further our understanding of Ras

91 in health and disease.

92

93

Results

95 **Structural analysis of experimental complex structures**

96 From the few experimentally determined complex structures of Ras with these RBDs, there are

97 some structural similarities that can be determined. The main interface on the site of Ras

98 consists of the two switch regions, switch 1 and switch 2. One of the main structural features

99   of the interface between Ras and RBDs is the formation of an intra-molecular β-sheet between

100   β2 on RAS and β2 on the RBD. This interaction is a highly conserved structural feature across

101   all available complex structures, with deviations in orientation of less than 1 Å (Figure 1B, see

102   Methods for MAE).

103   Analyzing the energetic profile of the interface *in silico* using the FoldX force field

104   shows that there are also some recurring hotspots in the interface (highlighted in Figure 1C).

105   I36, D38 and Y40 are well characterized as important residues for the interaction between Ras

106   and effector domains. Additionally, for the structures 3DDC and 1LFD, the mutation E31K was

107   introduced to stabilize the complex for crystallization. Our analysis confirms that this interface

108   mutation has indeed been favorable for the complex formation. Finally, the energetic

109   contributions of the function regions switch 1 and 2 to the interface were analyzed. Both relative

110   and absolute contributions are diverse, although switch 1 contributions to binding dominate

111   (Figure 1D). Altogether, the analysis of existing Ras-RBD effector structures indicates that

112   although there are many common features of the Ras RBD interface such as the intra-molecular

113   β-sheet or the hydrophobic patch around I36, the actual energetic contributions can come from

114   different parts of the interface. It also sets the basis for a successful homology modelling

115   approach.

116

117   Homology modelling and characterization of modelled interfaces

118   In order to study these interface features in a more diverse set of structures, homology models

119   of the complex between RAS and RBD were constructed for all proteins containing an RBD in

120   the human proteome.

121   The homology modelling pipeline is based on the already existing complex models

122   (Table S1). Also, with the recent release of AlphaFold2 (*14*) and the accompanying AlphaFold

123   Protein Structure Database (*15*), the RBD domains of all potential effectors were extracted from

124   that database and used. The structures of RBDs are predicted with good confidence by

125   AlphaFold2 and our analysis indicates that AlphaFold2 is reliable at predicting the RBD fold

126   (Figure S1). Additionally, AlphaFold2 complex modelling was attempted for all potential

127   complex structures and models which AlphaFold2 were confident in (by Predicted Alignment

128   Error (PAE)) and where the β-sheet alignment of the interface was within a tolerance to what

129   has been observed in crystal structures, were used as templates as well (Figure S2 and S3,

130   compare MAE Figure 1). There are two kinds of templates to use: 1) complex templates which

131   comprised of the already experimentally determined complex structures of Ras and RBDs, as

132   well as "good" AlphaFold2 predicted complex templates (Table S1), and 2) templates of the

133 RBDs alone which were extracted from the AlphaFold2 EMBL database (Table S2). An

134 overview over the pipeline is depicted in Figure 2. Homology modelling was performed using

135 homelette (*22*) with modeller and altMod. Evaluation of predicted structures was performed

136 using QMEAN, MolProbity and SOAP potentials. The top 300 models for each target for each

137 source of complex templates (experimental or AlphaFold2) were selected by combined score

138 and FoldX analysis (interaction energy and alanine scan) was performed.

139   In order to evaluate our approach, we generated a validation set, in which we created

140 models for the structures already solved by X-ray crystallography without using information

141 from the specific structure. Models generated in this validation set were compared to the

142 underlying ground truth by assessing their correlation of the *in silico* alanine scan results to

143 those of the crystal structures of interest. Using this ground truth, different methods to select

144 representative models from the hundreds of structures were assessed. In particular, we evaluated

145 the hyperparameters of an unsupervised learning pipeline comprised of different feature

146 selection and dimensionality reduction strategies followed by clustering with the OPTICS

147 algorithm (see Methods for more details about the hyperparameter space). After clustering,

148 three representative structures were chosen. Based on the performance in the validation set, the

149 optimal set of hyperparameters was chosen (Table S3, see Figures S4 and S5). The described

150 strategy for identifying representative structures was then applied to all target complexes and

151 we ended up with three representative complex structures for each effector.

152   Analogous to how we characterized the interfaces of the experimentally determined

153 complex structures before, we performed the same analysis on the complex models. The overall

154 FoldX interaction energies for the models are diverse, indicating that maybe some of the

155 complexes are energetically unfavorable and would not form (Figure S6B). In general, the

156 binding energies are lower than what would be observed in crystal structures, which is to be

157 expected. There are one or two outliers with regards to FoldX binding energy, namely RASSF8

158 and PIK3C2B. In particular, RASSF8 is also showing an uncommon hotspot profile, with

159 multiple unfavorable hotspots that are only appearing for this set of structures (Figure S6A).

160 Based on this behavior, RASSF8 is excluded from further analysis.

161   The analysis of hotspots confirmed the already established hotspots. I36, Y40, D38 and

162 E37 are the most commonly observed hotspots (Figure 3A, Figure S6A). Interestingly, while

163 I36, Y40, and E37 are exclusively favorable to the interaction with the effector protein, D38

164 seems to be also unfavorable in some of the structures (Figure 3B). The energetic diversity of

165 the hotspot D38 was further investigated in the models. For this, two models were picked for

166 which D38 was a favorable hotspot in the alanine scan analysis (Figure 3C: RASSF1, Figure

5

167   3D: RGL3), and two models were picked for which it was unfavorable (Figure 3E: ARAP2,

168   Figure 3F: RAPGEF3). Next, neighboring amino acids were analyzed. For favorable

169   interactions, we were able to observe positively charged amino acids. On RASSF1, we find

170   K216 and H217, whereas on RGL3, there is R630. These amino acids probably form strong

171   interactions with the negatively charged D38 on Ras. In contrast, for the models where D38

172   comes up as an unfavorable hotspot in the interface, we observe an uncharged, mostly

173   hydrophobic neighborhood.

174

175   **Estimation of binding energies**

176   One of the applications of the structural models that were generated was to use them for the

177   estimation of binding energies. Since experiments measuring the binding energy between two

178   proteins are experimentally very challenging and error-prone (*23*), we were implementing an *in*

179   *silico* approach. Also, while FoldX is good at predicting energy changes to interaction energy

180   or protein stability on mutation, the absolute interaction energies for protein complexes usually

181   are not well correlated with experimental values (*24*). Because of this, a supervised learning

182   regression pipeline was built based on features extracted from the modelled structures and a

183   collection of experimentally determined binding energies of different Ras-effector complexes

184   (Table S5). From a combination of different regressors, feature extraction procedures, and

185   hyperparameters, the best approach was determined using a cross validation strategy (see

186   methods). A support vector machine-based regressor (see hyperparameters in Table 4)

187   performed best in cross-validation with an R2 score of 0.53. Then, the performance of the best

188   approach was evaluated in an out-of-sample test set, where it achieved an R2 score of 0.77. The

189   model was then used to predict the interaction energies for the complexes without prior

190   experimental measurements (Figure 4, Table S5). The predicted binding affinities for our

191   models range 4 orders of magnitude, between the highest predicted affinity for BRAF of 0.02

192   μM to PIK3C2B with the lowest predicted affinity of 588 μM. The highest binding effectors

193   are quite well characterized (RAF family, PI3K, RASSF5, RIN1, RalGDS, AFDN (*8*)). As

194   experimental measurements of the PI3K family members are difficult as the RBD is not easy

195   to express and purify in isolation, it is noteworthy that we assign three of the good binding

196   affinities to PI3K family members (PIK3CA, PIK3CD, PIK3CG). A big group of effectors has

197   affinities in the range of 1 to 10 μM. For example, RASIP1 was previously in the "likely no

198   binding" category (*8*), and is now predicted to have an affinity in complex with Ras of 2.7 μM.

199   RAPGEF5 and RGL3 were previously in the "unknown" (*8*) category and have predicted

200   affinities of 9.2 and 5.6 μM, respectively. Another big group of effectors has affinities in the

201   range of 10 to 100 μM. Especially the first one could be interesting for modulation of binding

202   affinity by the piggy back mechanism (*7*).

203

204   **Switch contributions to binding**

205   Having Ras-effector structural models available allowed us to analyze the individual

206   contributions of switch 1 and switch 2 to binding using the results from alanine scanning

207   (similarly as done before for the X-ray Ras-effector structures). We find that generally most

208   structures are dominated by favorable switch 1 contributions (Figure 5 and Figure S7). Switch

209   2 contributions are surprisingly small. We also predict more contributions from regions outside

210   switch 1 and switch 2 as in the X-ray structures. For the two weak affinity binding groups (10

211   μM (K_d < 100 μM and (K_d > 100 μM) switch 1 contributions are in the range of 4 to 5 kcal/mol

212   with small (~1 kcal/mol) contributions from switch 2 and remaining parts involved in interface

213   formation. For the two strong affinity groups (K_d < 1 μM and 1 μM (K_d < 10 μM) switch 1

214   contributions increase to 6 to 9 kcal mol with also increasing switch 2 contributions (1 to 2

215   kcal/mol). We also observed negative switch contributions (mainly for switch 1), indicating

216   that these proteins are either not well predicted or non-binders. Indeed, all structures with

217   negative switch energies are weak binders.

218

219   **Branch pruning analysis using Ras-effector model structures**

220   Next, we were interested in exploring surface mutations on Ras that would selectively influence

221   the binding to some, but not all effectors. Both enhancing and inhibiting mutations are of

222   interest. This could enable the engineering of the Ras effector system to respond to stimuli in

223   different ways and to study selective sets of effectors. We previously reported a framework for

224   the identification and evaluation of so called 'branch pruning' mutations (*21*). Since our protein

225   is interacting with a many different effectors at the same time through the same interface it will

226   be quite unlikely to identify mutations that selectively target only one protein. Instead, it is more

227   likely that we will identify mutations that enhance in interaction with some proteins while

228   inhibiting some others.

229   Figure 6 shows a heatmap of all identified mutations of interest and their effect on all

230   effectors. Some interesting mutations to highlight are mutations around I36, that are almost

231   exclusively unfavorable while affecting almost every structure. D37 mutation are more

232   selective and also exclusively disruptive. D38 is mixed, as our analysis of the hotspot already

233   indicates. This is probably the best point to disrupt the system. Y40, interestingly, while being

234   a ubiquitous hotspot, is not a good spot for engineering the interface because mutations seem

7

235   to affect protein stability (compare with Figure S8A). Also of interest is that we detect both

236   mutations that increase and disrupt binding (Figure S8B).

237

Assessment of rewiring of Ras-effector interaction on a systems level

239   Finally, we want to evaluate the behavior of the Ras effector system based on our predicted

240   binding energies and based on the introduction of different branch pruning mutations. For this

241   means, we went back to our mathematical model of the Ras-effector system that incorporated

242   affinities and high-quality proteomics data in 29 human tissues (*7*). Here, all 29 tissue systems

243   were simulated at 20% and 90% GTP load on Ras. We are using exclusively the predicted

244   binding energies for this (except RASSF8, see methods). Overall, the results for the systems

245   without a branch pruning mutation are comparable to previous findings (Figure 7A; (*7*)). The

246   Raf family members ARAF, BRAF and RAF1 dominate the binding profile in complex with

247   Ras. Other effectors that are in high amount in complex with Ras in at least one of the 29 tissues

248   are RGL2, RASSF7, RASSF5, RASIP1, RALGDS, PIK3CD, PIK3CA, and AFDN.

249   Expanding on this, we introduced our branch pruning mutations to the mathematical

250   model. In total, there are 200 interface mutations that do not significantly affect overall protein

251   stability but affect binding to at least one of the effectors. Some of the branch pruning mutations

252   are able to dramatically change the system in all tissues, as can be seen from the example of

253   D38A (Figure 7B). With a single interface mutation, almost all RAF binding is quenched, and

254   other effectors start to compete for the binding. Interestingly, which effectors come up depends

255   on the tissue.

256   Next, we wanted to explore whether there are recurring states that the modelled system

257   assumes and whether these states are dependent on Ras-GTP load, the tissue, or interface

258   mutation. To this end, we applied uniform manifold approximation and projection (UMAP) to

259   our systems to transform a high-dimensional space of absolute and relative effector binding into

260   a 2D space. Then, we used OPTICS to identify areas of high density in this 2D plane and

261   assigned them into 19 distinct clusters (Figure 7C). For each cluster, we picked three of the

262   systems at random and visualized their relative effector binding (Figure S9). Each of these

263   clusters belongs to a different "state" that the Ras-effector system can be rewired into, with

264   systems belonging to a specific state showing similar trends in effector binding. Many of the

265   systems are dominated by ARAF binding, as it would be expected, but even for these there are

266   differences in the secondary effectors. To understand what the attributes of different "states" of

267   the systems are, two of them were picked and investigated for Ras-GTP load, tissue

268   composition, and interface mutation status (Figure 7DE). We find that there are different ways

269   a distinct state of the system can come together: for the state analyzed in Figure 7D, we can see

270   that it is composed of many different tissues, but only a handful of interface mutations, most of

271   them D38 mutations. This indicates that this state can be reached from many different tissues

272   by a specific, recurring set of mutations. In contrast, the state analyzed in Figure 7E is entirely

273   composed of a single tissue (lymph node) with many different mutations, indicated that this

274   state can only be reached by a specific tissue. Interestingly, both states analyzed are diverse in

275   terms of Ras-GTP load. To conclude, we identify 19 distinct states of the Ras-effector systems

276   and show that there are different mechanisms on how these states are formed.

277   Finally, we wanted to explore to which extend, for a specific effector, it is possible to

278   modulate its binding. For this, we visualized the possible changes to the relative amount of

279   effector bound to Ras across all systems tested (Figure S10). On the side of proteins that can be

280   negatively influenced, mostly the high-affinity binders such as RAF family proteins show up,

281   which is to be expected. Some effectors cannot be influenced by the branch pruning mutations,

282   either because they are energetically not affected or because their concentrations in any of the

283   29 tissues do not leave them in a position to compete for binding. Examples for this would be

284   RADIL or the TIAM family proteins. The effectors with the highest propensity to have their

285   binding enhanced are AFDN, RADIL, SNX27, RASSF5. In summary, based on a large set of

286   Ras-effector models, we predicted Ras branchegetic mutations and evaluate their binding in a

287   tissue-specific Ras competition model. The Ras mutations, once introduced into cells or tissues,

288   can be used as a tool to probe the contribution of specific effector pathways to an output or

289   cellular phenotype.

290

291

Discussion

292

293   In this work we have shown a complete structural reconstruction of Ras and the RBDs of its

294   effectors based on state-of-the-art technology. We used these structural models to investigate

295   the workings of the interface between Ras and its effectors, as well as to search for and identify

296   potential branch pruning mutations on Ras that would alter the behavior of the underlying

297   system. We analyzed the effects on the steady state of the Ras effector system.

298   Recent advances in structural modelling, mostly by the development and release of

299   AlphaFold2 and similar algorithms have pushed the field of structural bioinformatics forward.

300   This development was crucial for the quality of this study. While a normal homology modelling

301   pipeline would have been successful, the inclusion of both aspects of AlphaFold2 models was

302   essential for the quality of the results. On the one hand, finding additional potential complex

303   templates diversified the possible configurations of the interface that we could use to generate

304   our models. On the other hand, having high quality templates of the RBDs enabled us to

305   improve the structural predictions. Interestingly, with all the advancements that AlphaFold2

306   brought, this combined AlphaFold2/homology modelling approach yielded more consistent and

307   better results for this particular question.

308   There is a growing body of literature about how Ras dimerizes or forms multimers, or

309   interacts with the membrane to modulate the signaling. All these aspects have been deliberately

310   left out for this approach. The essence of the interaction of Ras and its effectors is the binary

311   protein-protein interaction between the Ras switch regions and the RBD. This common feature

312   was the focus of this work, and we believe that other factors such as dimer/multimerization of

313   Ras, the composition of the membrane, etc., are only modulators for this interaction.

314   By creating a complete structural system, we were able to investigate and understand

315   the interactions of Ras with its effector molecules on a different level. Crucially, it enabled us

316   to analyze how *in silico* mutations of the system could affect its behavior. This is an interesting

317   approach for a lot of different systems, not just the Ras effector system. However, there are

318   certainly challenges. The prediction and verification of a protein-protein interface can be very

319   complicated, and the techniques for modelling protein-protein interactions are not sufficiently

320   developed to easily translate that approach to a larger scale. Probably the main reason why it

321   was possible to construct this structural system for the Ras-effector system was because it is a

322   conserved domain-domain interaction between homologue proteins. Although the sequence

323   identity of the RBD sequences is not well preserved anymore, the structural fold of these

324   domains has been preserved. Additionally, there is also the preserved mode of binding by the

325   formation of the intermolecular β-sheet. These factors were favorable for the construction of

326   the structural system and would need to be addressed if this approach were to be taken to a

327   higher scale. Recent publications are already able to work on system-wide structural prediction

328   for interactions (*25*). These approaches are very promising for the structural characterization of

329   known protein complexes and can identify high-confidence novel interactions as well.

330   However, for the exhaustive characterization of a full system, especially with transient

331   interactions, the distinction between true and false positives seems to remain challenging.

332   Finally, we hope that our structural and systems analysis of the branch pruning interface

333   mutations will enable interesting experimental setups that study different downstream pathways

334   from Ras. Some of the more promising candidates are AFDN, RADIL, SNX27, RASSF5. The

335   downstream pathways of Ras that are best characterized are the RAF-MEK-ERK signaling

336   pathway and the PI3K-AKT signaling pathway. However, some of the other proteins might

337   play a role in a physiological or pathophysiological context as well. AFDN is essential for the

338   organization of adhesion between cells (*26*), function that is often impaired in cancer (*27*).

339   RADIL is also linked to cell adhesion, and recent data showed that knockdown of was linked

340   to decreased cell proliferation and invasion (*28*). SNX27 is also part of signaling pathways that

341   link to cell adhesion and barrier function (*29*). RASSF5 is a tumor suppressor and has been

342   shown to inhibit growth and invasion and to induce apoptosis (*30*). Importantly, all branchegetic

343   mutations were studied on a systems level using a tissues specific Ras competition model. These

344   models can easily be adapted for specific cell systems of interest, provided that estimates for

345   Ras and effector abundances are available. Altogether, this work contributed to a quantitative

346   understanding of a key cellular hub protein – Ras.

347

348

349   STAR★Methods

350   Detailed methods are provided in the online version of this paper and include the following:

351   • **KEY RESOURCE TABLE**

352   • **RESOURCE AVAILABILITY**

353

354   **Lead contact**

355   Further information and requests for resources should be directed to Philipp Junk

356   (philipp.junk@ucdconnect.ie)

357

358   **Code availability**

359   The code will soon be deposited on Zenodo.

360

361   METHOD DETAILS

362   **Experimentally determined Ras-effector complex structures**

363   Structures were downloaded from the PDB. In the case where multiple models were available,

364   the best one by MolProbity score was selected (*31*) (*32*). The PDB files were processed so that

365   all GTP and Mg2+ annotations were in the expected format. The list of used template structures

366   can be found in Table S1.

367

368   **Interface characterization**

369 Hotspots residues were determined by FoldX *in silico* Alanine scan (*33*) (*24*) (*20*) (*34*) (*35*). In

370 detail, each residue on both RAS and the effector was mutated to alanine and the *in silico* change

371 of binding energy ΔΔG was determined as such:

$$\Delta\Delta G\ alascan = \Delta G(mut) - \Delta G(wt)$$

373 A positive ΔΔG value indicates that the mutated residue is involved favorably in the interaction,

374 whereas a negative ΔΔG value indicates that the mutation to alanine improved the interaction

375 between the two proteins. The standard error for ΔΔG values in FoldX is around +- 0.8 kcal/mol.

376 In order to identify the most important residues for the respective interaction, a ΔΔG cut-off of

377 1.6 kcal/mol was chosen for the investigation of crystal structures. Since the energies are

378 systematically lower in the modelled complex structures, a cut-off of 1.2 kcal/mol was chosen

379 for those.

380 Based on the alanine scan results, the contribution of the two major functional regions

381 in the RAS interface (*36*), switch 1 (residues 20-42) and switch 2 (residues 56-76) to the

382 interaction were determined. For each functional region and the remainder of Ras, the ΔΔG

383 values were filtered by abs(ΔΔG) > 0.8 kcal/mol and then summed up. The definition of the

384 switch regions used here is more generous than what is normally used in the literature, and it

385 would be probably more correct to label them as "switch-influenced" regions. These residue

386 ranges aim to capture the regions in the interface that are affected by the movement of switch

387 1 and 2 during the transition from the inactive GDP bound state to the active GTP bound state.

388 The conserved intra-molecular β-sheet between the β2 sheet on Ras with the β2 sheet on the

389 effector is evaluated by measuring the differences in inter-molecular distances between the β

390 sheets in the experimentally solved complex structures and comparing them to the ones in a

391 structure of interest. This measurement has been named Measured Alignment Error (MAE) in

392 this manuscript:

$$MAE\left(res_{eff}, res_{ras}\right) = \frac{\sum_{i=1}^{n} \sqrt{(dist_{model} - dist_{ref})^2}}{n}$$

394 with $dist_{model}$ being the Euclidean distance between two residues in the model of interest;

395 $dist_{ref}$ being the Euclidean distance between the residue on Ras and the corresponding residue

396 in the reference structure, as determined by structural alignment using TMalign (*37*); and n

397 being the number of reference structures. The references used were the X-ray complex

398 templates found in the PDB (Table S1). When the MAE is calculated for X-ray complex

399 templates, the comparison with itself is removed from the calculation.

400

401

AlphaFold2 determined single structures for RBDs

402

403 For each protein containing one or more potential RBD sequences as identified in (*6*), the full

404 structure was obtained from the AlphaFold Protein Structure database (*14*) (*15*) (Release 1,

405 accessed September 2021). The sequences of the RBDs were obtained from Pfam (*38*) and

406 UNIPROT (*39*). The part of the structures that correspond to RBD sequences was subsequently

407 extracted from the AlphaFold2 structures. The list of used templates as well as information

408 about the extracted RBD domains can be found in Table S2.

409

410 **AlphaFold2 determined complex structures: generation and selection**

411 AlphaFold2-multimer/ColabFold (version 1.3.0) was run on a local computer (*40*) (*14*) (*41*).

412 Multiple sequence alignments were generated from MS2seq (*42*) (*43*) (*44*) (*45*). Complex

413 models were generated with HRAS as interaction partner. For each target, five models were

414 generated with additional template search and five without additional template search (*40*) (*46*).

415 The resulting model were relaxed as per ColabFold defaults (*40*) (*47*). For each model, several

416 metrices were evaluated. Firstly, AlphaFold2's Predicted Alignment Error (PAE) was taken

417 into consideration. PAE is an estimation of the error of pairwise distances between residues,

418 that can be used to assess how confident AlphaFold2 is in the inter-domain arrangement of its

419 models. Secondly, FoldX (*33*) (*24*) (*34*) (*35*) interaction energies were determined for the

420 complexes as described above. Finally, the expected orientation of the inter-molecular β sheets

421 was evaluated by MAE. The best model by MAE was selected for each target, and then all

422 complex structures with a MAE > 1A were removed.

423

424 **Homology modelling pipeline**

425 Alignment generation, model generation and model evaluation were performed using the

426 homelette homology modelling interface (*22*). Inputs to the homology modelling pipeline were

427 complex structures of Ras in complex with some effectors, either experimentally determined or

428 selected from AlphaFold2 complex predictions, as well as AlphaFold2 models of all RBDs of

429 interest. For each target, all combinations of the RBD template with all complex templates were

430 used to generate 300 models of the target in complex with HRAS each. The different sources

431 for the complex templates were run separately with slight differences in the modelling

432 procedure. For complex templates of experimental origin, TMalign (*37*) was used to generate a

433 structure-based sequence alignment based on the RBD in the complex template and the single

434 RBD of the target structure. Then, with those two templates as inputs, models were generated.

13

435 For complex templates of *in silico* origin, structure-based sequence alignments were generated

436 with TMalign (*37*) as described. As an additional template a HRAS single structure (5P21 (*3*))

437 was used in the modelling process since the AlphaFold2-generated complex templates, in

438 contrast to the complex templates of experimental origin, do not have information about the

439 important heteroatoms GTP and MG2+ in the Ras part of the structure. Models were generated

440 using modeller (*48*) (*49*) with the altMOD extension (*50*). All models generated were evaluated

441 using QMEAN (*51*) (*52*), MolProbity, (*31*) (*32*) and SOAP (*53*) potentials. A combined score

442 was determined based on borda count as such:

$$combined\ score(X) = \sum_{i=1}^{m} n - rank_i(X)$$

444 For an observation X with i…m being a collection of evaluation criteria and n the total number

445 of observations. A structure with high borda score is a structure that performs well across all

446 metrices.

447 For each of the different sources of complex templates (experimental or *in silico*), 300 models

448 were selected in a first selection step based on the combined score. These 600 models per target

449 were then further analyzed using FoldX. *In silico* interaction energies were determined and *in*

450 *silico* alanine scan was performed (see Interface Characterization).

451 In addition to generating models for unknown targets, a set of validation models based

452 on the experimentally solved models were generated as well. For each of the seven PDB

453 complex structures, 300 models were generated. Inputs were restricted so that the structure to

454 be modelled would not be used as a complex template, but only the remaining six

455 experimentally derived complex templates. For each validation target, 300 models were

456 selected as described above.

457 Using the results from the analysis with FoldX, representative structures were selected

458 based on an unsupervised learning workflow. As the clustering algorithm of choice, Ordering

459 Points To Identify the Clustering Structure (OPTICS) (*54*) was used, as implemented in scikit-

460 learn (https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html). OPTICS is a density-based

461 clustering method, which unlike the more popular k-means clustering does not require a

462 manually set input of the number of clusters. Additionally, OPTICS is able to label data points

463 as outliers. Several approaches to feature selection and/or feature engineering, hyperparameters

464 of OPTICS, as well as methods for selecting representative structures from clusters were

465 evaluated against the set of validation models (Table 3). To select the best combination of

466 hyperparameters, for each combination, the *in silico* alanine scan results of the representative

467 structures were correlated to the results of corresponding PDB structures, and the combination

14

468 of hyperparameters with the most stable performance across all validation sets (by minimum z-
469 score of the correlation against the PDB structure for all 7 validation sets) was chosen. The best
470 combination of hyperparameters is highlighted in (Table 3).
471

472 **Estimation of binding energies**

473 Supervised learning based on a number of FoldX-derived features was used in order to estimate
474 binding energy of complex models. The features consisted of the FoldX interaction energy, the
475 energy contribution of switch 1, 2 and the remainder of the Ras protein interface (see Interface
476 characterization), and the $\Delta\Delta G$ values for hotspot residues on Ras (cut off 1.2 kcal/mol). All
477 features were standardized and highly intercorrelated features were removed. Data preparation
478 was performed in R.

479   The experimental measurements of the dissociation constant between Ras-effector
480 complexes were collected from several publications (see Table S5, also available as
481 supplementary data). Effectors that were experimentally determined as non-binders were
482 removed from the data set due to uncertainty how to encode them with the varying technical
483 limitations on detectable binding energies at the time of their publication. Also, the models
484 generated for RASSF3 seem to be outliers with regards to FoldX interaction energy (see figure
485 7) and were therefore removed from the prediction. A test set was manually chosen from the
486 available experimental measurements to cover the full spectrum of experimentally determined
487 interaction energies. At the end, this gave us a training set with 51 observations (17 * 3 models)
488 and a testing set with 12 observations (4 * 3 models).

489   Supervised   learning   was   performed   in   scikit-learn
490 (https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html) using different combinations of
491 feature selection algorithms and regressors. Feature selection was performed using either f-
492 regression or mutual information as implemented by scikit-learn. Regression was evaluated for
493 different algorithms with various hyperparameter spaces (see Table S4). All combination of
494 feature selection and regression were evaluated using Group-K-Fold cross validation within the
495 training data, with k=5 and the groups corresponding to the three structural models chosen for
496 each target. Models were evaluated using R2 score. The best performing combination was
497 trained on the whole training data and evaluated against the test data. Finally, this model was
498 used to predict the binding energies for the complex structures without experimental data.
499

500 **Branch pruning**

15

501 FoldX was used to evaluate the effect of interface mutation on the binding energies in the

502 modelled complex structures. As previously described, (21) the FoldX command PSSM was

503 used to evaluate the changes in binding energy, and the FoldX command PosScan was used to

504 evaluate if mutations impacted the stability of RAS.

505 All mutations that were destabilizing HRAS in either of two structures (3TGP and 5P21)

506 were removed from the analysis (cutoff 1.6 kcal/mol). Then, between the three models for each

507 target structures, it was checked if a mutation had a noticeable impact ($> +-$ 1kcal/mol) in at

508 least two of the three structures. If so, the changes in binding energy for all models above/below

509 the cutoff were averaged.

510

511 **Systems analysis**

512 A mathematical model of the Ras-effector system was set up as previously described (6) (7).

513 The model is based on classic ligand-receptor kinetics according to the assumption of

514 conservation of mass. A system of ordinary differential equations was set up and steady states

515 were calculated as described. The reactions are expressed as such:

516 $R + E_i \xrightarrow{(k_i, k_{-i})} RE_i$ With R representing the molar concentration of Ras, E_i the molar

517 concentration of an effector of the set of $i \in (1, 2, ..., 54)$ effectors (all modelled proteins,

518 except for proteins from the Ubiquitin family) and RE_i the molar concentration of a Ras-effector

519 complexes. The complex is formed at rate k_i and dismantled at rate k_{-i}. These rates define the

520 dissociation constant:

521
$$K_{d_i} = \frac{k_{-i}}{k_i}$$

522 Due to the assumption of mass conservation,

523
$$R_{tot} = R + \sum_{1}^{54} RE_i$$

524 and

525
$$E_{tot_i} = E_i + RE_i$$

526 the system to solve therefore is:

527
$$R_{tot} = R + \sum_{1}^{54} RE_i$$

528
$$E_{tot_i} = E_i + RE_i$$

529
$$K_{d_i} = \frac{(R * E_i)}{RE_i}$$

16

530   For the set of $i = \{1,2, ...,54\}$ and can be numerically solved for RE_i. The system was solved

531   using SciPy.

532   The concentrations for the species in the model were taken from the supplementary data

533   of (*7*), in which molar concentrations were derived from high-quality proteomics data set of 29

534   different human tissues (*55*). The concentration of Ras proteins (HRAS, NRAS, KRAS) were

535   pooled together and then multiplied with a loading factor to take into account the balance

536   between active (GTP bound) and inactive (GDP bound) Ras. This loading factor was set to 0.2

537   for a normal RAS system, and 0.9 for a system hyperactivated by an oncogenic hotspot

538   mutation.

539   The binding affinities were taken from the predicted binding affinities (see Estimation

540   of binding affinities), with the exception of RASSF3 for which we are not confident in our

541   structural predictions. An experimentally determined binding affinity for RASSF3 was

542   substituted from (*56*). Additional sets of binding affinities based on the branch pruning analysis

543   were evaluated as well. For this, the predicted binding affinities were adapted by the ddG values

544   from the branch pruning analysis. For the case that a system with hyperactivated Ras was

545   considered, it was made sure that common oncogenic hotspots such as G12, G13 and Q61 do

546   not influence the binding energies for any effector.

547   Some sets of parameters were not solvable in a physically meaningful solution space

548   (negative concentrations) or at all, these parameter sets were removed from further analysis.

549   To gain an overview over all solved system, absolute and relative effector bindings were

550   transformed using UMAP (*57*) and visualized. OPTICS clustering (*54*) was performed on the

551   UMAP transformed data (parameters: min_sample = 25, min_cluster_size=200).

552

Data analysis and visualization

553

554   All data analysis, unless otherwise noted was performed in R using the tidyverse environment

555   (https://www.R-project.org/). Visualizations were generated using ggplot2 (H. Wickham.

556   ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.).

557   Visualizations of protein structures were generated using PyMol (version 2.5.0)

558   (https://pymol.org/2/support.html).

559

Figure legends

Figure 1. Noticeable features of the Ras-RBD interfaces

(A) Overview of interface features for the Ras-RAF1 interface. Highlighted are the intramolecular β-sheet alignment, the assembly of the Ras interface by the switch regions and energetic hotspots in the interface.

(B) β-sheet alignment for experimentally determined complexes (MAE). A lower value indicates a more similar alignment of the intramolecular β-sheets compared to other crystal structures.

(C) FoldX alanine scan hotspots on RAS for experimentally determined complex structures. The color scale is confined to the limits [-1.6, 1.6]. Positive ΔΔG values indicate a loss in binding energy upon mutation, which reflects that the respective amino acid contributes to binding.

(D) Energetic contributions of functional regions based on alanine scan analysis.

Figure 2. Overview over the homology modelling pipeline

The principal homology modelling steps are shown from left to right with the steps of "Target Identification", "Template Identification", "Alignment Generation", "Complex Model Generation", "Evaluation & Selection 1", and "Evaluation & Selection 2". In the "Validation" process, the homology modelling pipeline is run on a set of 7 RBD domains from known x-ray structures. The "Production" process describes the generation of 61 domains (RBD and Ubiquitin super-fold domains as controls).

Figure 3. Energy hotspots in modelled structures

(A) Heatmap of FoldX alanine scan averaged from the three representative structures on each target. The color scale is confined to the limits [-1.6, 1.6]. Hotspot residues with a ΔΔG >= 1.2 or <= -1.2 kcal/mol were marked.

C-F) Local neighborhood of D38 in structures where D38 is a favorable hotspot (C: RASSF1, D: RGL3) or an unfavorable hotspot (E: ARAP2, F: RAPGEF3). Ras and the effector structures are visualized in blue and green, respectively. Polar interactions between charged amino acids are indicated with dashes.

Figure 4. Results of the affinity prediction for all Ras-effector complex structures

18

592   Visualization of predicted binding affinities. The three representative structures for each target

593   are visualized as black dots, with the averaged affinity (based on averaged energy) is visualized

594   as a red cross.

595

596   **Figure 5. Switch contributions for the summed-up energy contributions grouped by their**

597   **predicted binding affinity**

598   (A) Energy contributions were separately calculated for switch 1 switch or the rest of the Ras

599   interface by summing up energy energies from the *in silico* alanine scan analysis. Complexes

600   were grouped based on their predicted binding affinities into four groups.

601   (B – E) Examples of the switch contributions for each of the four groups. Visualized are BRAF

602   (panel B), AFDN_1 (panel C), ARHGAP20 (panel D), and PIK3C2B (panel E).

603

604

605   **Figure 6. Energetic characterization of RAS interface mutations that affect effector**

606   **binding**

607   The color scale is confined to the limits [-3.2, 3.2]. Hotspot residues with a $\Delta\Delta G >= 1$ or $<= -1$

608   were marked.

609

610   **Figure 7. Ras-effector interaction rewiring at a systems level**

611   (A-B) Effects of interface mutation on the Ras-effector system. Heatmap of effectors bound to

612   Ras at 20% Ras-GTP load in 29 tissues in (panel A) WT interface and in (panel B) with the

613   effects of a D38A mutation.

614   (C-E) Overview over the possible states the Ras-effector system can assume. UMAP

615   transformation of all simulated systems. Similar systems were identified using OPTICS

616   clustering (panel C). 19 identified clusters are colored in rainbow colors, with systems classified

617   as outliers colored in grey. Characterization of two identified state clusters in terms of Ras-GTP

618   load, tissues, interface mutations, and most representative effectors (panels D and E). For the

619   tissue and mutation pie charts, all groups that were smaller than 5% of the total observations

620   were collapsed into the "Other" group.

621

622

623

19

624   (C) UMAP transformation of all simulated systems. Similar systems were identified using

625   OPTICS clustering. 19 identified clusters are colored in rainbow colors, with systems classified

626   as outliers colored in grey.

627   (D) and (E) Characterization of two identified state clusters in terms of Ras-GTP load, tissues,

628   interface mutations, and most representative effectors. For the tissue and mutation pie charts,

629   all groups that were smaller than 5% of the total observations were collapsed into the "Other"

630   group.

631

632

633

634

635

636

637

638

639

640

641

1.   R. Gasper, F. Wittinghofer, The Ras switch in structural and historical perspective. *Biol Chem* **401**, 143-163 (2019).

2.   E. F. Pai *et al.*, Structure of the guanine-nucleotide-binding domain of the Ha-ras oncogene product p21 in the triphosphate conformation. *Nature* **341**, 209-214 (1989).

3.   E. F. Pai *et al.*, Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 A resolution: implications for the mechanism of GTP hydrolysis. *Embo j* **9**, 2351-2359 (1990).

4.   C. Kiel, L. Serrano, The ubiquitin domain superfold: structure-based sequence alignments and characterization of binding epitopes. *J Mol Biol* **355**, 821-844 (2006).

5.   D. K. Simanshu, D. V. Nissley, F. McCormick, RAS Proteins and Their Regulators in Human Disease. *Cell* **170**, 17-33 (2017).

6.   V. Ibáñez Gaspar, S. Catozzi, C. Ternet, P. J. Luthert, C. Kiel, Analysis of Ras-effector interaction competition in large intestine and colorectal cancer context. *Small GTPases* **12**, 209-225 (2021).

7.   S. Catozzi, M. Halasz, C. Kiel, Predicted 'wiring landscape' of Ras-effector interactions in 29 human tissues. *NPJ Syst Biol Appl* **7**, 10 (2021).

8.   C. Kiel, D. Matallanas, W. Kolch, The Ins and Outs of RAS Effector Complexes. *Biomolecules* **11**, (2021).

9.   S. Catozzi *et al.*, Reconstruction and analysis of a large-scale binary Ras-effector signaling network. *Cell Commun Signal* **20**, 24 (2022).

10.   C. Kiel, E. Verschueren, J. S. Yang, L. Serrano, Integration of protein abundance and structure data reveals competition in the ErbB signaling network. *Sci Signal* **6**, ra109 (2013).

11.   B. N. Kholodenko, J. B. Hoek, H. V. Westerhoff, Why cytoplasmic signalling proteins should be recruited to cell membranes. *Trends Cell Biol* **10**, 173-178 (2000).

12.   C. Kiel, M. Foglierini, N. Kuemmerer, P. Beltrao, L. Serrano, A genome-wide Ras-effector interaction network. *J Mol Biol* **370**, 1020-1032 (2007).

13.   C. Kiel, L. Serrano, Prediction of Ras-effector interactions using position energy matrices. *Bioinformatics* **23**, 2226-2230 (2007).

14.   J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).

15.   M. Varadi *et al.*, AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50**, D439-d444 (2022).

16.   R. C. Gimple, X. Wang, RAS: Striking at the Core of the Oncogenic Circuitry. *Front Oncol* **9**, 965 (2019).

17.   A. R. Moore, S. C. Rosenberg, F. McCormick, S. Malek, RAS-targeted therapies: is the undruggable drugged? *Nat Rev Drug Discov* **19**, 533-552 (2020).

18.   A. G. Stephen, D. Esposito, R. K. Bagni, F. McCormick, Dragging ras back in the ring. *Cancer Cell* **25**, 272-281 (2014).

19.   Q. Zhong *et al.*, Edgetic perturbation models of human inherited disorders. *Mol Syst Biol* **5**, 321 (2009).

20.   C. Kiel, L. Serrano, Structure-energy-based predictions and network modelling of RASopathy and cancer missense mutations. *Mol Syst Biol* **10**, 727 (2014).

21.   P. Junk, C. Kiel, Engineering of Biological Pathways: Complex Formation and Signal Transduction. *Methods Mol Biol* **2315**, 59-70 (2021).

22.   P. Junk, C. Kiel, HOMELETTE: a unified interface to homology modelling software. *Bioinformatics* **38**, 1749-1751 (2021).

23.   I. Jarmoskaite, I. AlSadhan, P. P. Vaidyanathan, D. Herschlag, How to measure and evaluate binding affinities. *Elife* **9**, (2020).

24.   J. Schymkowitz *et al.*, The FoldX web server: an online force field. *Nucleic Acids Res* **33**, W382-388 (2005).

692   25.   I. R. Humphreys *et al.*, Computed structures of core eukaryotic protein complexes. *Science*
693   **374**, eabm4805 (2021).
694   26.   A. F. Citalán-Madrid, A. García-Ponce, H. Vargas-Robles, A. Betanzos, M. Schnoor, Small
695   GTPases of the Ras superfamily regulate intestinal epithelial homeostasis and barrier
696   function via common and unique mechanisms. *Tissue Barriers* **1**, e26938 (2013).
697   27.   T. A. Martin, W. G. Jiang, Loss of tight junction barrier function and its role in cancer
698   metastasis. *Biochim Biophys Acta* **1788**, 872-891 (2009).
699   28.   B. H. Choi *et al.*, Identification of Radil as a Ras binding partner and putative activator. *J Biol*
700   *Chem* **296**, 100314 (2021).
701   29.   S. P. Zimmerman, C. L. Hueschen, D. Malide, S. L. Milgram, M. P. Playford, Sorting nexin 27
702   (SNX27) associates with zonula occludens-2 (ZO-2) and modulates the epithelial tight
703   junction. *Biochem J* **455**, 95-106 (2013).
704   30.   X. H. Zhou *et al.*, RASSF5 inhibits growth and invasion and induces apoptosis in osteosarcoma
705   cells through activation of MST1/LATS1 signaling. *Oncol Rep* **32**, 1505-1512 (2014).
706   31.   V. B. Chen *et al.*, MolProbity: all-atom structure validation for macromolecular
707   crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).
708   32.   C. J. Williams *et al.*, MolProbity: More and better reference data for improved all-atom
709   structure validation. *Protein Sci* **27**, 293-315 (2018).
710   33.   R. Guerois, J. E. Nielsen, L. Serrano, Predicting changes in the stability of proteins and protein
711   complexes: a study of more than 1000 mutations. *J Mol Biol* **320**, 369-387 (2002).
712   34.   C. Kiel, L. Serrano, C. Herrmann, A detailed thermodynamic analysis of ras/effector complex
713   interfaces. *J Mol Biol* **340**, 1039-1058 (2004).
714   35.   C. Kiel, L. Serrano, Cell type-specific importance of ras-c-raf complex association rate
715   constants for MAPK signaling. *Sci Signal* **2**, ra38 (2009).
716   36.   I. R. Vetter, A. Wittinghofer, The guanine nucleotide-binding switch in three dimensions.
717   *Science* **294**, 1299-1304 (2001).
718   37.   Y. Zhang, J. Skolnick, TM-align: a protein structure alignment algorithm based on the TM-
719   score. *Nucleic Acids Res* **33**, 2302-2309 (2005).
720   38.   J. Mistry *et al.*, Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412-d419
721   (2021).
722   39.   UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**, D480-d489
723   (2021).
724   40.   M. Mirdita *et al.*, ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679-
725   682 (2022).
726   41.   R. Evans *et al.*, (2022).
727   42.   M. Mirdita, M. Steinegger, J. Söding, MMseqs2 desktop and local web server app for fast,
728   interactive sequence searches. *Bioinformatics* **35**, 2856-2858 (2019).
729   43.   M. Mirdita *et al.*, Uniclust databases of clustered and deeply annotated protein sequences
730   and alignments. *Nucleic Acids Res* **45**, D170-d176 (2017).
731   44.   A. L. Mitchell *et al.*, MGnify: the microbiome analysis resource in 2020. *Nucleic Acids Res* **48**,
732   D570-d578 (2020).
733   45.   M. Steinegger *et al.*, HH-suite3 for fast remote homology detection and deep protein
734   annotation. *BMC Bioinformatics* **20**, 473 (2019).
735   46.   H. Berman, K. Henrick, H. Nakamura, Announcing the worldwide Protein Data Bank. *Nat*
736   *Struct Biol* **10**, 980 (2003).
737   47.   P. Eastman *et al.*, OpenMM 7: Rapid development of high performance algorithms for
738   molecular dynamics. *PLoS Comput Biol* **13**, e1005659 (2017).
739   48.   A. Sali, T. L. Blundell, Comparative protein modelling by satisfaction of spatial restraints. *J*
740   *Mol Biol* **234**, 779-815 (1993).
741   49.   B. Webb, A. Sali, Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc*
742   *Bioinformatics* **54**, 5.6.1-5.6.37 (2016).

743   50.   G. Janson *et al.*, Revisiting the "satisfaction of spatial restraints" approach of MODELLER for
744   protein homology modeling. *PLoS Comput Biol* **15**, e1007219 (2019).
745   51.   P. Benkert, S. C. Tosatto, D. Schomburg, QMEAN: A comprehensive scoring function for
746   model quality assessment. *Proteins* **71**, 261-277 (2008).
747   52.   P. Benkert, M. Biasini, T. Schwede, Toward the estimation of the absolute quality of
748   individual protein structure models. *Bioinformatics* **27**, 343-350 (2011).
749   53.   G. Q. Dong, H. Fan, D. Schneidman-Duhovny, B. Webb, A. Sali, Optimized atomic statistical
750   potentials: assessment of protein interfaces and loops. *Bioinformatics* **29**, 3158-3166 (2013).
751   54.   M. Ankerst, M. M. Breunig, H.-P. Kriegel, J. Sander, Optics. *ACM SIGMOD Record* **28**, 49-60
752   (1999).
753   55.   D. Wang *et al.*, A deep proteome and transcriptome abundance atlas of 29 healthy human
754   tissues. *Mol Syst Biol* **15**, e8503 (2019).
755   56.   S. Rezaei Adariani *et al.*, A comprehensive analysis of RAS-effector interactions reveals
756   interaction hotspots and new binding partners. *J Biol Chem* **296**, 100626 (2021).
757   57.   M. W. Dorrity, L. M. Saunders, C. Queitsch, S. Fields, C. Trapnell, Dimensionality reduction by
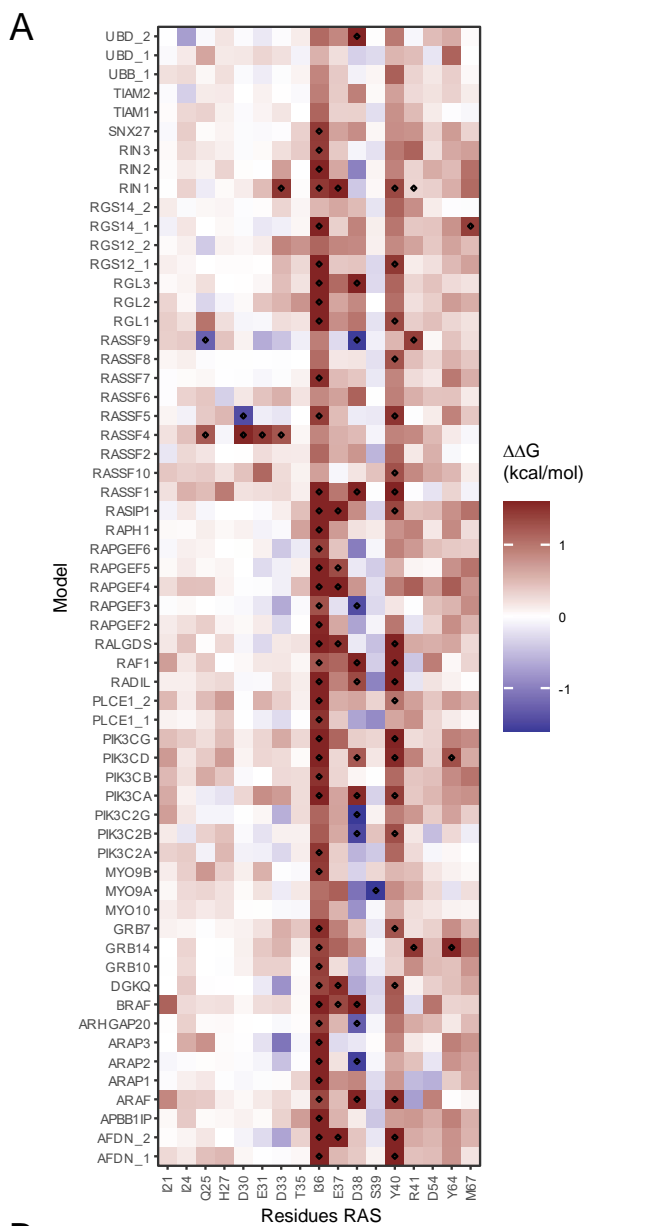758   UMAP to visualize physical and genetic interactions. *Nat Commun* **11**, 1537 (2020).

759

Figure 1
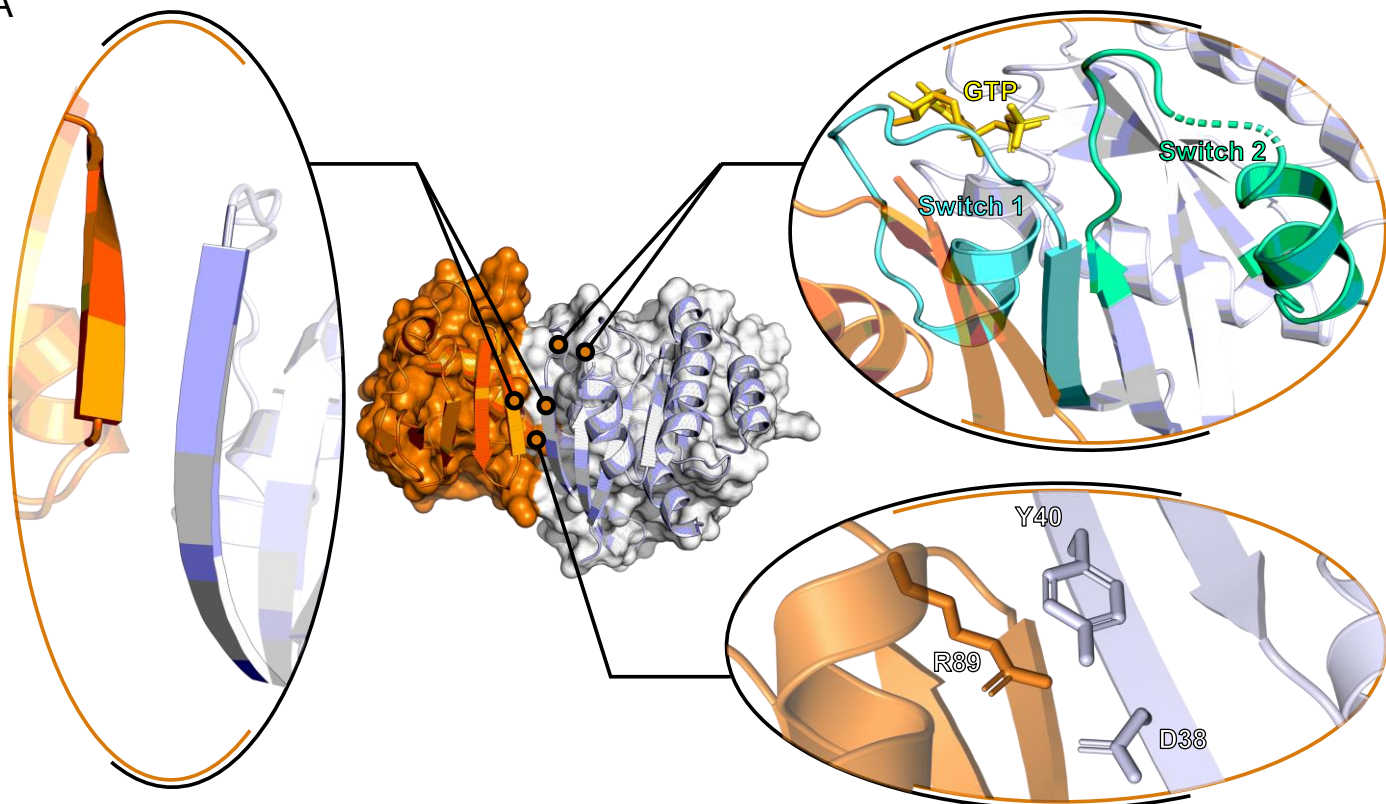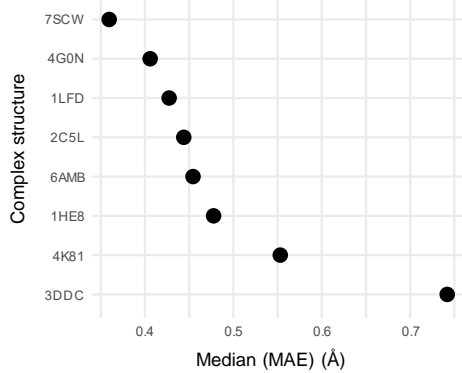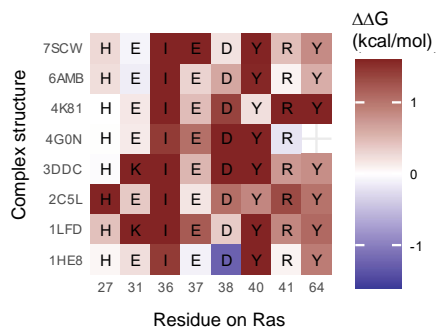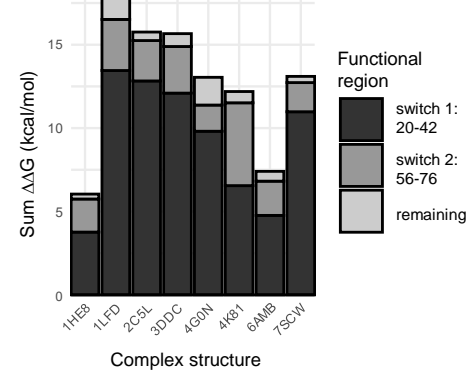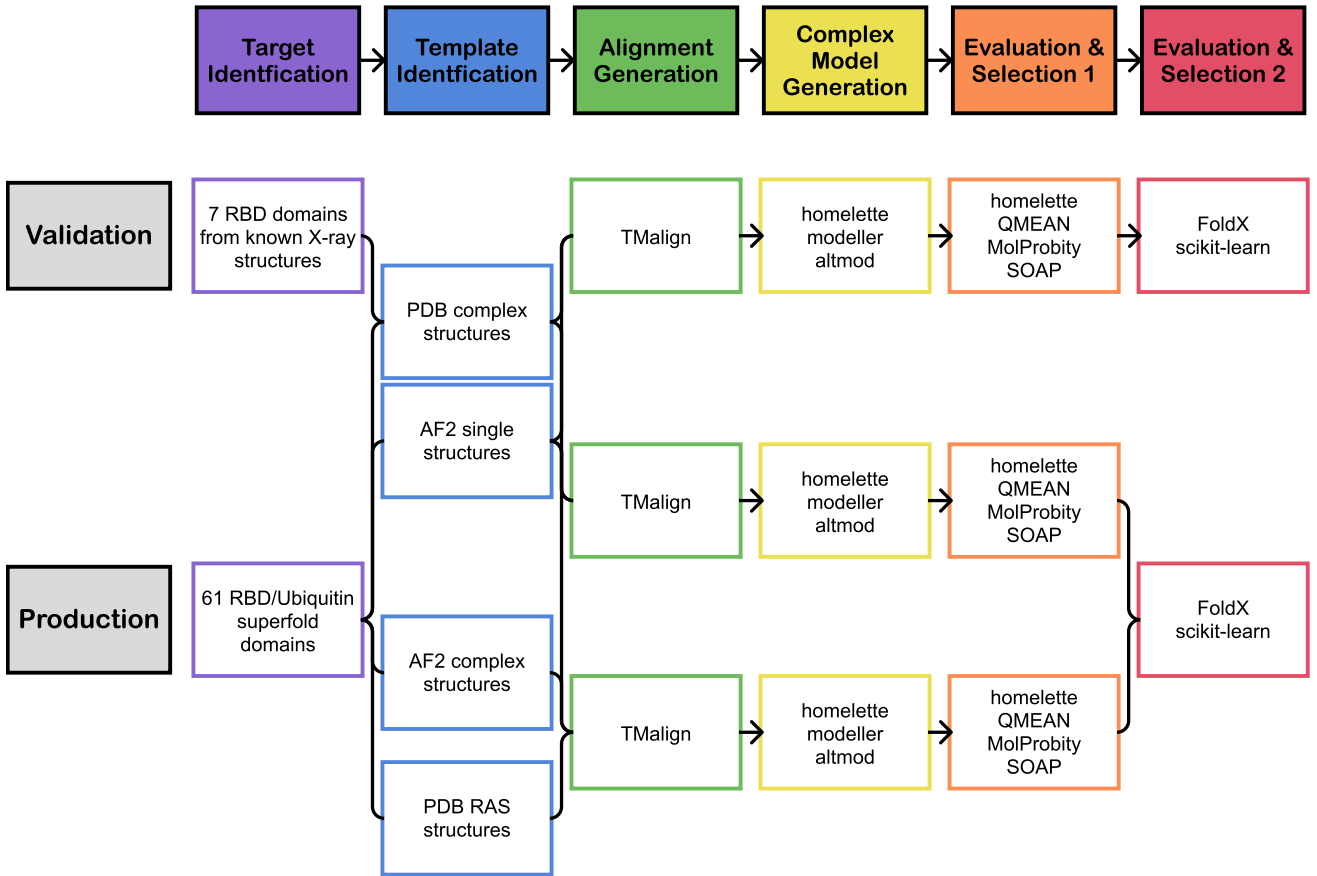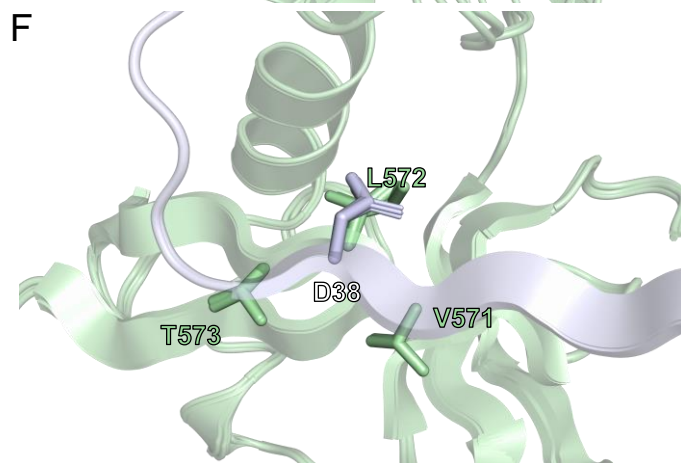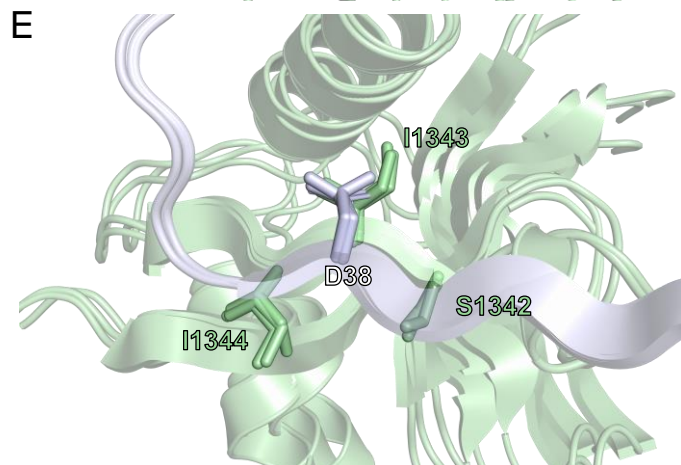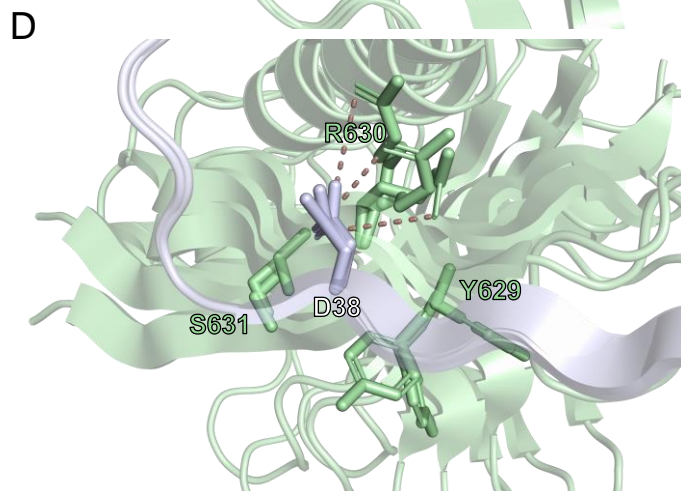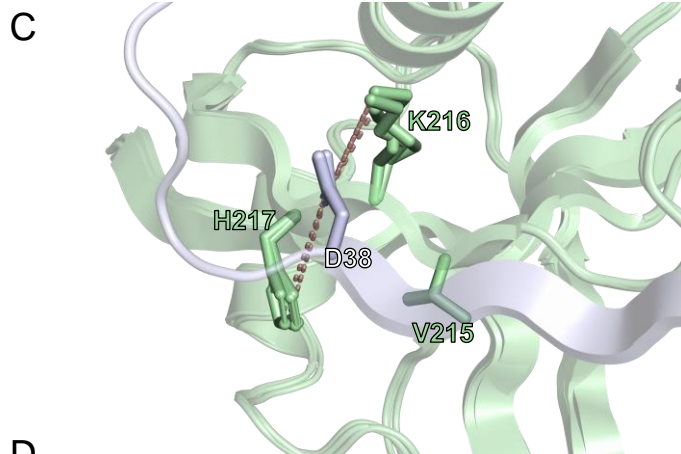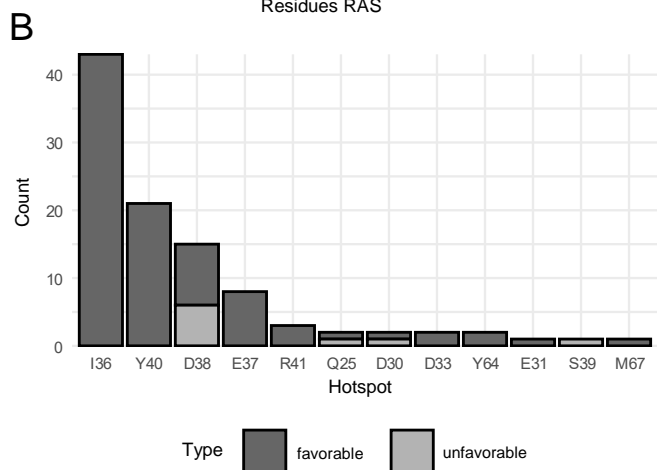
Figure 2

Figure 3

Figure 4
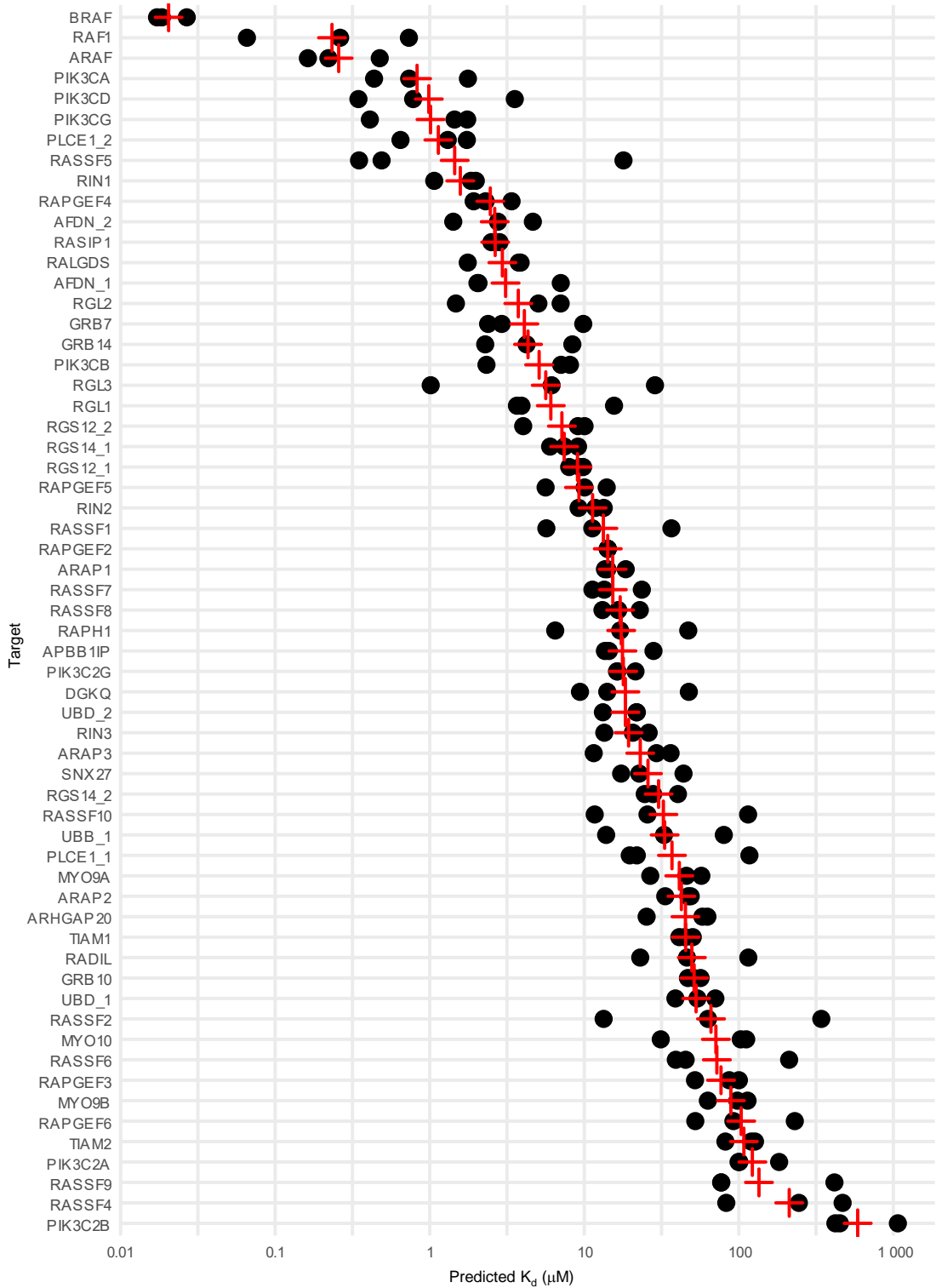
Figure 5

Figure 6

Figure 7