1
2
3
4
5
6
7

Anti-citrullinated protein antibodies arise during affinity maturation of germline antibodies to carbamylated proteins in rheumatoid arthritis

10
11
12
13

Marta Escarra-Senmarti[1][¥], Michael Chungyoun[2][¥], Dylan Ferris[1], Jeffrey J. Gray[2]*, Felipe Andrade[1]*

Affiliations:

[1]Division of Rheumatology, The Johns Hopkins University School of Medicine. Baltimore, MD, 21224. USA.

[2]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, 21287, USA.

¥ These authors contributed equally to this work

*Corresponding author.

Jeffrey J. Gray
3300 N Charles St
Maryland Hall, Room 208
Baltimore Maryland 21218
☎ 410-516-5313
✉ jgray@jhu.edu

Felipe Andrade
5200 Eastern Avenue, Center Tower
Mason Lord Building, Suite 6000, Room 608
Baltimore Maryland 21224
☎ 410-550-8665
✉ andrade@jhmi.edu

49 **Abstract**

50 Why autoantibodies in rheumatoid arthritis (RA) primarily target physiologically modified proteins,

51 called citrullinated proteins, is unknown. Recognizing the inciting event in the production of anti-

52 citrullinated protein antibodies (ACPAs) may shed light on the origin of RA. Here, we demonstrate

53 that ACPAs originate from germline-encoded antibodies targeting a distinct but structurally similar

54 modification, called carbamylation, which is pathogenic and environmentally driven. The transition

55 from anti-carbamylated protein (anti-CarP) antibodies to ACPAs results from somatic

56 hypermutations, indicating that the change in reactivity is acquired via antigen-driven affinity

57 maturation. During this process, a single germline anti-CarP antibody transitions from anti-CarP

58 to double positive (anti-CarP/ACPA) to ACPA according to the pattern and number of somatic

59 hypermutations, explaining their coexistence and diverse specificity in RA. Artificial intelligence-

60 based structural modeling revealed that an ACPA and its germline precursor exhibit distinct

61 structural and biophysical properties, and pointed to heavy-chain tryptophan 48 (H-W48) as a

62 critical residue in the differential recognition of citrullinated vs. carbamylated proteins. Indeed, a

63 single methionine substitution in H-W48 changes the antibody specificity from ACPA to anti-CarP.

64 These data indicate that the existence of germline-encoded anti-CarP antibodies is most likely

65 the first event in the production of ACPAs during the early stages of RA development.

66

Introduction

The production of antibodies to modified self-antigens is a hallmark in rheumatoid arthritis (RA), an autoimmune disease of unknown etiology characterized by immune-mediated damage to synovial joints [1]. Antibodies to citrullinated and carbamylated proteins are of particular interest in RA because of their high prevalence, specificity, association with erosive disease, and appearance during the preclinical phase of RA [2-6]. Citrullination is a physiological posttranslational modification (PTM) best known for its role in gene regulation and skin keratinization [7]. During this process, arginine residues are converted to citrulline by the activity of the peptidylarginine deiminase (PAD) enzymes (**Fig. 1a,** upper panel) [8]. Carbamylation is a chemical modification that results from the addition of a carbamoyl moiety to free functional groups of proteins, peptides, and free amino acids [9-11]. This process is driven by environmental factors (such as smoke from biomass burning, biofuel usage, cooking, and tobacco smoking) and during inflammation, and mainly results from the reaction of isocyanic acid with a primary amine or a free sulfhydryl group [2,9-11]. When carbamylation occurs on the ε-amino group of lysine residues, it generates homocitrulline residues (**Fig. 1a**, lower panel) [9-12].

Citrullination and carbamylation have in common that they neutralize the positive charge of arginine and lysine, respectively, creating modified residues with similar structure, charge, and hydrophobicity (i.e., citrulline and homocitrulline, respectively). The only difference is that homocitrulline is one methylene group longer than citrulline **(Fig. 1a)** [2]. Nevertheless, although citrullination and carbamylation generate nearly identical amino acid residues, these PTMs have significant differences in biology and potentially for RA [2]. Citrullination is a physiological process across multiple tissues [13] and therefore, citrullinated proteins are poorly immunogenic *in vivo* [14]. This is explained because, at least in mice, T and B cells reactive to citrullinated proteins are negatively selected during development [15-17]. Indeed, the role of citrullinated proteins to induce arthritis in mice is controversial [18-20] and the antibody response is usually not citrulline-specific [14,21,22]. In contrast, carbamylation is a pathological protein modification generated in conditions

93   associated with oxidative stress [12,23-26]. It is considered in the group of PTM derived products

94   (PTMDPs), which includes other protein modifications such as oxidation, glycation, glycoxidation,

95   lipoxidation, and carbonylation [27,28]. PTMDPs are irreversible, accumulate throughout life, and are

96   distinguished by the spontaneous binding to free reactive groups of proteins, which has a

97   significant impact on protein structure and biological functions, especially in the context of aging,

98   physiological stress, and inflammation [24,27,28].

99   Notably, different to citrullinated proteins, carbamylated proteins are highly immunogenic

100   [14,29] and do not require a xenoreactive response to induce antibodies. Indeed, similar to other

101   chemical modifications in proteins [30], anti-carbamylated protein (anti-CarP) antibodies are

102   efficiently generated by immunization with autologous carbamylated proteins [20,31,32]. Thus,

103   although citrulline and homocitrulline are almost identical, the immune system can distinguish

104   between the two PTMs under normal conditions. This is likely explained because different to

105   citrullination, which is physiological, carbamylation is environmentally driven and thereby absent

106   during B cell development. Thus, B cells able to react with carbamylated proteins are not

107   subjected to negative selection. Interestingly, immunization with carbamylated peptides or

108   autologous carbamylated albumin, but not with citrullinated peptides/albumin, renders mice

109   susceptible to an aggressive form of erosive arthritis following a subsequent intra-articular

110   injection of citrullinated peptides [20]. Moreover, anti-CarP antibodies precede disease onset in

111   monkeys with collagen-induced arthritis [33]. Taken together, these data suggest a pathogenic link

112   between carbamylation and citrullination in the induction of arthritis.

113   Although a potential mechanistic connection between anti-CarP antibodies and ACPAs

114   remains unknown, the study of RA patient-derived monoclonal ACPAs supports that binding to

115   carbamylated sequences results from cross-reactivity with ACPAs [34]. Further understanding of

116   the relationship between ACPAs and anti-CarP antibodies, however, has been hampered by the

117   search for additional PTMs cross-reactive with ACPAs (e.g., acetyl-lysine and ornithine) [35,36],

118   resulting in greater confusion about the origin of these antibodies [2].

119   Here, we provide a simple and rational explanation for the mechanistic link between

120   citrullination and carbamylation in RA, integrating critical differences in their origin (physiologic vs.

121   pathologic) and immunogenicity (tolerized vs. non-tolerized), which may explain why citrullinated

122   proteins are abnormally targeted by the immune system in RA. Our studies indicate that antibody

123   binding to carbamylated proteins is germline-encoded. B cells expressing germline anti-CarP

124   antibodies escape negative selection, most likely because carbamylation is not physiologic, which

125   explains why carbamylated proteins are normally immunogenic. However, since citrulline and

126   homocitrulline are nearly identical, the data also indicate that germline-encoded anti-CarP

127   antibodies can generate a mixture of anti-CarP, anti-CarP/ACPA and ACPAs as the result of

128   antigen-driven affinity maturation, most likely in the context of genetic susceptibility, tobacco

129   smoking, inflammation and hypercitrullination, among other factors. Overall, these findings imply

130   that the presence of germline-encoded anti-CarP antibodies is the initial step in the production of

131   disease-specific antibodies during disease initiation in RA.

132

133   **Results**

134   **Somatic hypermutations dictate ACPA specificity for citrullinated or carbamylated**

135   **cellular substrates.** E4 is an RA patient-derived monoclonal ACPA that has been extensively

136   characterized, including crystal structure analysis bound to citrullinated antigens and its effect on

137   collagen antibody-induced arthritis (CAIA) in mice [37-39]. While the specificity of E4 has been tested

138   against synthetic citrullinated peptides [38], its reactivity to citrullinated and carbamylated cellular

139   substrates has not been addressed. To gain insights into the reactivity of E4 to citrullinated and

140   carbamylated cellular substrates and the role of somatic mutations in antigen binding, we

141   generated recombinant antibodies in ExpiCHO cells using the E4 available sequence [37,38], as well

142   as a version reverted to germline (both heavy and light chains). As substrates, 293T cells mock

143   or transfected to co-express PAD2 and PAD4 were stimulated with ionomycin to generate control

144   and hypercitrullinated cells, respectively. In addition, 293T cells were incubated with 1M

145   potassium cyanate to generate carbamylated proteins (**Fig. 1b**). In contrast to lysates from control

146   and carbamylated cells, E4 showed prominent binding to hypercitrullinated cells (**Fig. 1c,** left

147   panel), demonstrating that citrullinated proteins are the major target of E4. When E4 was reverted

148   to its germline form (E4GL), however, citrullination was no longer detected (**Fig. 1c,** right panel),

149   indicating that reactivity to citrullinated proteins by E4 is gained as result of antigen-driven affinity

150   maturation. Unexpectedly, while E4 showed poor recognition of carbamylated proteins (**Fig. 1c**,

151   left panel), E4GL showed prominent binding to carbamylated proteins (**Fig. 1c**, right panel),

152   implying that carbamylation is a primary target of E4GL. Moreover, the data indicate that, while

153   somatic hypermutations promote E4 binding to citrullinated proteins, they have a significant

154   impact on E4 targeting carbamylated proteins. Thus, unlike the current paradigm, these findings

155   suggest that cross-reactivity plays a minimal role in the binding of ACPAs to citrullinated or

156   carbamylated proteins. Instead, the data supports that the stage of antibody maturation seems to

157   define the ACPA specificity to citrullinated or carbamylated proteins.

158

159   **The Chai-1 all-atom protein structure prediction method is capable of modeling the**

160   **interactions of E4 with a citrullinated peptide.** To understand the molecular basis that

161   determines the distinct specificity between E4 and E4GL, we used structural modeling. Artificial

162   intelligence (AI) methods for all-atom biomolecular structure prediction are capable of modeling

163   the joint structure of complexes including proteins, nucleic acids, small molecules, ions, and

164   modified residues [40-44]. However, although these AI methods demonstrate substantially improved

165   accuracy over many specialized tools for protein-ligand interactions, a comprehensive evaluation

166   of their performance on specific antibody-antigen complexes involving non-canonical amino acids

167   (NCAAs, for instance, citrulline and homocitrulline) remains unexplored. As no crystal structure of

168   E4GL bound to a carbamylated peptide is available, but the structure of E4 bound to the

169   citrullinated type II collagen peptide CII-Cit-13 (PDB: 5OCX) has been resolved [38], we first aimed

170   to identify the AI model best suited for predicting the E4GL-carbamylated peptide complex. To

171 achieve this, we benchmarked several all-atom biomolecular structure prediction methods,

172 namely Chai-1 [40], AlphaFold3 [41], Boltz-1 [44], RoseTTAFold-AA [42], and Protenix [43].

173 Among the five structure prediction methods evaluated, only Chai-1, AlphaFold3, and

174 Boltz-1 successfully positioned the citrulline residue of CII-Cit-13 within the paratope pocket of E4

175 (**Supplementary Fig. 1** and **Supplementary Table 1**). Both Chai-1 and AlphaFold3 predicted

176 complexes and accurately depicted the three hydrogen bonds between E4 heavy (H) chain

177 variable domain and citrulline that mediate its recognition. Two bonds involve the oxygen atom of

178 the ureido moiety in citrulline, which interacts with the hydroxyl group of H-S51 and the amine

179 nitrogen of H-W48, and a third bond between the ureido group nitrogen and the carboxamide

180 oxygen of H-N105 [38]. However, the Boltz-1 predicted complex lacked the crucial interaction

181 between H-S51 and H-W48. Out of the eight regions of the E4 variable domain [i.e., six

182 complementarity-determining region (CDR) loops, heavy chain (HC) framework, and light chain

183 (LC) framework], Chai-1 outperformed AlphaFold3 in predicting the conformations of four regions:

184 CDR H3, CDR L1, and both the HC and LC framework regions. Both methods performed equally

185 well in predicting the conformation of CDR H1. Because Chai-1 accurately positioned citrulline

186 within the E4 pocket and outperformed all other methods in predicting most of the E4 CDR loop

187 and framework conformations—including the correct placement of H-W48, a crucial residue for

188 citrulline binding in the HC framework [38,39]—we chose this method to predict the conformation of

189 E4GL with citrullinated and carbamylated peptides. Indeed, Chai-1 has been previously found to

190 outperform AlphaFold3 at general protein-ligand complex prediction, and antibody-protein

191 complex prediction [40].

192

193 **The differing sequence, structural, and biophysical properties of E4 and E4GL**

194 **determine their respective antigen proclivity.** To gain further insights into the role of somatic

195 hypermutations in the transition from germline-encoded anti-CarP antibodies to ACPA, we

196 investigated the differences in heavy-chain variable/light-chain variable (VH/VL) packing, charge,

197 and solvent accessible surface area (SASA) when bound with either a citrullinated or

198 carbamylated peptide. We predict the complex of E4 with CII-Cit-13 to represent an ACPA, and

199 E4GL with a peptide sequence from carbamylated α1-antitrypsin (A1AT-HCit), which is targeted

200 by anti-CarP antibodies and found in RA synovial fluid [45], to represent an anti-CarP antibody. We

201 also predicted the complex of E4 with A1AT-HCit and E4GL with CII-Cit-13 to determine whether

202 we observe the same binding preferences computationally as we did experimentally.

203 　　　The VH/VL packing angle, distance, and opening angle of E4 and E4GL are similar in

204 value (**Supplementary Table 2**), leading to a generally similar variable domain when overlapped

205 with each other (**Fig. 2a**). However, E4GL displays a noticeably larger VH/VL opposite opening

206 angle (**Supplementary Table 2**), which may explain the increased width of the antibody paratope

207 pocket (**Fig. 2b**), as well as its preference for homocitrulline, which is one methylene group longer

208 than citrulline (**Fig. 1a**). The CDR loops of E4 and E4GL antibodies differ in surface area, charge,

209 and aromatic content. E4 generally has larger SASA values, particularly in H1, L1, and L2,

210 indicating more exposed loops (**Supplementary Table 3**). E4 is more negatively charged in loops

211 H1 and L1, while E4GL is often neutral or positively charged. E4GL also contains aromatic

212 residues in loops like H2, where E4 lacks them (**Supplementary Table 3**). Despite these

213 differences, both antibodies have similar charge and aromatic content in H3 (**Supplementary**

214 **Table 3**). **Fig. 2c** further highlights the differences in charge distribution, with E4 exhibiting more

215 negative charge on the HC side and the citrulline/homocitrulline (Cit/HCit) pocket, while E4GL has

216 stronger positive charge areas on the LC side. The observed biophysical differences between E4

217 and E4GL, including variations in surface area, charge distribution, and loop structure, are key

218 factors influencing their distinct antigen binding preferences. These results reinforce prior findings

219 that somatic mutations in antibodies lead to significant biophysical changes compared to their

220 germline versions, influencing their antigen binding properties and potentially driving their distinct

221 specificity profiles [46].

222

E4 and E4GL display opposing peptide binding characteristics. We initially predicted the structure of E4 and E4GL in complex with CII-Cit-13 or A1AT-HCit to exhaustively investigate the binding preferences of E4 and E4GL. The predicted complex of E4 with CII-Cit-13 shows key interactions between citrulline and three critical E4 HC residues [38] – H-W48, H-S51, and H-N105 (**Fig. 3a**). In contrast, the predicted E4-A1AT-HCit complex does not position homocitrulline within the antibody pocket (**Fig. 3b**). This difference in binding is also reflected in the predicted binding affinity, where E4 exhibits a stronger binding affinity to CII-Cit-13 than A1AT-HCit (**Supplementary Table 4**). While citrulline can still enter the E4GL pocket (**Fig. 3c**), structural differences between E4 and E4GL are apparent. In E4GL, the oxygen group of H-N105 points upward, whereas in E4 it points toward the center of the pocket. Additionally, two hydrogen bonds are lost in E4GL: the oxygen group of H-N105 no longer forms an intramolecular bond with the ureido group nitrogen, and H-W48 is too far from citrulline to interact (**Fig. 3c**). However, in the E4GL-A1AT-HCit complex, homocitrulline forms hydrogen bonds with H-N105 and H-S51 (**Fig. 3d**), with H-S51 oriented downward in both E4GL complexes, unlike the upward orientation observed in E4. The structural changes in E4GL shifts its predicted peptide affinity as well, where E4GL displays a stronger binding affinity to A1AT-HCit than to CII-Cit-13 (**Supplementary Table 4**). The structural differences between E4 and E4GL, particularly the orientation of H-N105 and H-S51 side chains, influence their peptide binding. Importantly, only H-W48 interacts with citrulline in E4, while in all other complexes, H-W48 is not involved. These differences in structure and interactions likely underlie the distinct binding preferences observed experimentally for E4 and E4GL, as well as the importance of H-W48 in citrulline but not homocitrulline binding.

H-W48 is a key residue that dictates ACPA binding to citrulline or homocitrulline residues. While H-W48 is essential for citrulline binding by E4 (**Fig. 3a** and [38,39]), it is worth noting that this residue is germline encoded in all human IGHV genes (except IGHV3-64 and IGHV3-64D) as either H-W47, H-W48 or H-W49. This implies that somatic mutations in residues other

249   than H-W48 cause structural changes in E4 that positions H-W48 in the correct orientation to

250   interact with citrulline. Indeed, when H-W48 was changed to methionine in E4 (E4M), the antibody

251   lacked reactivity to citrullinated proteins (**Fig. 3e**), as previously described [39]. However, it was

252   surprising to find that, while E4M lacked its ACPA specificity, it regained reactivity to carbamylated

253   proteins (**Fig. 3e**), mimicking the germline antibody E4GL (**Fig. 1c**, right panel). Moreover,

254   consistent with the structural and experimental data, full-length carbamylated A1AT was only

255   detected by E4GL and E4M, but not E4 (**Fig. 3f**). Collectively, the data imply that while H-W48 is

256   essential for citrulline recognition by E4, this residue interferes with the efficient binding of E4 to

257   homocitrulline. Thus, the specificity to citrulline or homocitrulline appears to be defined by H-W48,

258   with its position determined by the pattern of somatic hypermutations. Notably, further analysis of

259   published RA patient-derived monoclonal ACPAs [37,38,47,48] showed that although these antibodies

260   are highly mutated, H-W47/48/49 is conserved (i.e., unmutated) among ACPAs (**Fig. 3g**), implying

261   that like E4, the presence of tryptophan around position H-48 may also determine the specificity

262   to citrulline or homocitrulline in other ACPAs.

263   To gain further insights into the origin of ACPAs, we produced three additional RA-derived

264   monoclonals for which the VH and VL gene sequences are publicly available and have been

265   shown to have strong reactivity to citrullinated peptides and some citrullinated proteins (RA3, A2-

266   2 and F12-2) [48]. Unexpectedly, except for a couple of proteins detected in hypercitrullinated cells,

267   monoclonals A2-2 and F12-2 mainly recognized carbamylated proteins (**Supplementary Fig. 2a**

268   and **2b,** respectively), suggesting that these are primarily anti-CarP antibodies. Nevertheless,

269   when A2-2 and F12-2 were reverted to their germline forms (A2-2GL and F12-2GL), they

270   remained anti-CarP, albeit with less reactivity (**Supplementary Fig. 2a** and **2b,** respectively),

271   indicating that, like E4, they are derived from mutated germline antibodies to carbamylated

272   proteins. In the case of RA3, it showed reactivity to both citrullinated and carbamylated proteins

273   (Cit/CarP) (**Fig. 3h,** left panel) and consistent with E4, germline reverted RA3 (RA3GL) was only

274 reactive to carbamylated proteins (**Fig. 3h,** middle panel) and a single mutation in H-W47 to

275 methionine (RA3M) made the antibody anti-CarP specific (**Fig. 3h,** right panel).

276

277 **Single germline-encoded anti-CarP antibodies generate anti-CarP, ACPAs and**

278 **double reactive anti-CarP/ACPAs as result of affinity maturation.** Despite that RA-derived

279 monoclonal antibodies catalogued as ACPAs (i.e., E4, RA3, A2-2 and F12-2) have distinct

280 reactivity to citrullinated, carbamylated or both Cit/CarP, it is intriguing that, regardless of their

281 specificity, they all derive from hypermutated germline precursors reactive to carbamylated

282 proteins, implying a common mechanistic origin. To further understand the ontogenic relationship

283 between anti-CarP, ACPAs and anti-CarP/ACPAs, we analyzed two sets of RA-derived

284 monoclonal ACPAs that originate from single germline precursors. The first group includes four

285 phylogenetically related ACPAs for which the VH and VL gene sequences are publicly available

286 (termed T2H12, T2E04, T3G09 and T2D09) [47]. In addition, we generated their common germline

287 precursor (termed Gr1GL), including the putative germline-encoded CDR3 (**Fig. 4a**) [47]. The

288 second group includes five related monoclonal ACPAs (C3-4, G7-6, C1-2, A3-6 and B3-6)[48],

289 which also share a common germline precursor (termed Gr2GL) (**Supplementary Fig. 3**).

290 Consistent with the specificity of other germline reverted ACPAs (i.e., E4GL, A2-2GL, F12-

291 2GL, and RA3GL), Gr1GL and Gr2GL only targeted carbamylated proteins but not control or

292 hypercitrullinated cell lysates (**Fig. 4a** and **Supplementary Fig. 3a**), supporting the notion that

293 ACPAs originate from germline-encoded anti-CarP antibody precursors. Interestingly, ACPAs

294 derived from Gr1GL and Gr2GL showed different antigen specificities, generating antibodies with

295 preferential binding to carbamylated proteins (e.g., T2E04), citrullinated proteins (e.g., T2D09 and

296 C3-4), and with dual Cit/CarP reactivity (e.g., T2H12, T3G09, G7-6, C1-2, A3-6 and B3-6) (**Fig.**

297 **4a** and **Supplementary Fig. 3a**). Moreover, it is remarkable that the pattern of citrullinated and/or

298 carbamylated proteins detected by each antibody is different, indicating that these antibodies

299 have preference for distinct modified proteins, even when derived from the same germline

300   precursor. The finding that somatic hypermutations in a single germline-encoded anti-CarP

301   antibody can generate a mixture of anti-CarP, anti-CarP/ACPA and ACPA with distinct

302   specificities explains their variable coexistence and diverse specificity among patients with RA.

303   Although Gr1GL-derived ACPAs that gained reactivity to citrullinated proteins (i.e., T3G09

304   and T2D09) had a higher number of missense mutations (**Fig. 4b** and **4c**), suggesting a transition

305   from anti-CarP to ACPA according to the number of somatic hypermutations, this rule was not

306   observed in Gr2GL-derived ACPAs, in which antibodies with a lower number of mutations (i.e.,

307   C3-4 and G7-6) (**Supplementary Fig. 3b** and **3c**) showed better reactivity to citrullinated proteins

308   (**Supplementary Fig. 3a**). Thus, the data suggest that somatic hypermutations in germline-

309   encoded anti-CarP antibodies can favor either specificity (ACPA, anti-CarP and/or anti-

310   CarP/ACPA), which is likely determined by the type of antigen that dominates during the process

311   of affinity maturation.

312   To further address the role of H-W48 (H-W49 in ACPAs derived from Gr1GL) in ACPA

313   specificity, we mutated the tryptophan to methionine in three antibodies with the highest reactivity

314   to citrullinated proteins (T3G09 to T3G09M, T2D09 to T2D09M, and G7-6 to G7-6M). Consistent

315   with other ACPAs, antibodies T3G09M, T2D09M and G7-6M lacked reactivity to citrullinated

316   proteins (**Fig. 4d** and **Supplementary Fig. 3d**), supporting that H-W48/49 is critical for binding to

317   citrullinated proteins in all tested ACPAs. However, only antibodies T3G09M and G7-6M, not

318   T2D09M, were strongly reactive to carbamylated proteins (**Fig. 4d** and **Supplementary Fig. 3d**),

319   suggesting that in some ACPAs, other residues than H-W/47/48/49 are involved in their lack of

320   reactivity to carbamylated proteins.

321

322   **Protein acetylation is a cross-reactive antigen for some ACPAs, but it is not a primary**

323   **target of germline-encoded anti-CarP antibodies.**

324   Because acetylated-lysine residues resemble homocitrulline residues (except for the side-

325   chain terminal amine, which is replaced by a methyl group) (**Fig. 5a**), and some ACPAs can be

326 cross-reactive to acetylated proteins [34,49], it has been hypothesized that this modification may be

327 relevant to the induction of ACPAs in RA. Using acetylated BSA (**Fig. 5b**), we found that all

328 antibodies derived from E4 (E4GL, E4 and E4M) and RA3M showed reactivity to this antigen (**Fig.**

329 **5c** and **Fig. 5d**), indicating that different to homocitrulline and citrulline in proteins, acetylated-

330 lysine residues are simply cross-reactive with no specificity driven by the E4 germline sequence,

331 somatic hypermutations or H-W47/48/49. Notably, no other of the 12 ACPAs, including their

332 germline and mutated H-W47/48/49 forms (except for RA3M), showed reactivity to acetylated

333 BSA (**Fig. 5c** and **5d**). Thus, while acetylated-lysine residues can be cross-reactive to some

334 ACPAs, which could be relevant for RA pathogenesis [2], our data do not support that this

335 modification is involved in the origin of ACPAs in RA.

336

337 **Discussion**

338 The analysis of serum samples collected years before clinical onset of RA has been critical

339 in recognizing that the presence of disease-specific antibodies is one of the first events in the

340 initiation of RA [1,5,6]. Furthermore, the study of RA patient-derived monoclonal ACPAs has

341 demonstrated that they are mostly generated from non-citrulline reactive precursors via antigen-

342 driven affinity maturation [50,51]. In this study, we showed that carbamylated proteins are the target

343 of a set of ACPAs reverted to their germline form, implying that anti-CarP antibodies are

344 precursors that can lead to the production of ACPAs during disease initiation in RA.

345 Depending on the pattern and number of somatic hypermutations, the data indicate that

346 germline anti-CarP antibodies can generate distinct antibody subsets against carbamylated or

347 citrullinated proteins, as well as double reactive antibodies. Because this combination of

348 antibodies is most likely generated at random from different anti-CarP B cell precursors during

349 the course of RA, it explains why studying the serologic appearance of anti-CarP antibodies and

350 ACPAs during the preclinical and clinical phases of the disease has been insufficient to address

351 the sequence of events leading to the origin of ACPAs. Instead, the analysis of autoantibodies in

352 their germline form has been a useful strategy for understanding their antigenic origin and how

353 somatic hypermutations create pathogenic antibodies, some of which exhibit multiple

354 autoreactivities [52,53].

355 Interestingly, although autoreactive germline antibodies can be precursors of pathogenic

356 autoantibodies [52,53], germline-encoded anti-CarP antibodies do not apply to this model because

357 they are not properly autoreactive. Instead, they are directed to a novel protein structure, which

358 is generated in the context of inflammation and noxious environmental stimuli [2]. Thus, the

359 presence of germline antibodies to carbamylated proteins is not explained by a defect in central

360 tolerance. Rather, since carbamylated proteins are expected to be absent during normal

361 development, B cells with immunoglobulin gene rearrangements reactive to carbamylated

362 proteins would not be eliminated and would remain in the normal B cell repertoire in some healthy

363 individuals.

364 Several RA-associated components are likely to concur to convert a non-pathogenic anti-

365 CarP B cell into an ACPA producing cell. Chronic exposure to environmental factors that promote

366 protein carbamylation, like tobacco smoking, may initiate the activation of naïve anti-CarP B cells

367 [54,55]. Similarly, the abnormal production of carbamylated proteins by leukocyte-derived

368 peroxidases as result of chronic inflammation, such as at mucosal sites, is another potential

369 source of carbamylated antigens [25,56-58]. While these components can initiate an immune

370 response to carbamylated proteins, this process is most likely self-limited, as it is with any foreign

371 antigen under normal conditions. However, given that homocitrulline and citrulline are almost

372 identical, the concurrence with hypercitrullination and RA-associated genetic susceptibility is likely

373 to promote the further expansion and affinity maturation of anti-CarP B cells into ACPA B cells by

374 targeting abnormally citrullinated proteins [2]. For instance, microbial-induced hypercitrullination is

375 an abundant source of citrullinated neoantigens [59,60], which may expose citrulline residues on

376 structural epitopes recognized by anti-CarP antibodies.

377   Notably, this study demonstrates the power of AI-based all-atom protein structure

378   prediction in modeling antibody-peptide interactions, particularly for antibodies with NCAAs. Our

379   results show that Chai-1 effectively modeled the binding of E4 to citrullinated peptides,

380   outperforming other tools, particularly in predicting CDR loop and framework conformations.

381   Additionally, structural differences between E4 and E4GL, including changes in VH/VL packing,

382   charge distribution, and aromatic content, explain their distinct binding preferences for citrullinated

383   and carbamylated peptides. These findings underscore the importance of structural details like

384   the orientation of key residues, such as H-W48, in determining antigen specificity.

385   Our work represents a significant step forward in using AI methods for modeling antibody-

386   NCAA interactions, offering a novel approach for studying autoimmune diseases and therapeutic

387   antibody design. By demonstrating the accuracy of Chai-1 in predicting antibody binding to

388   modified peptides, we establish a foundation for more efficient antibody characterization in the

389   absence of crystal structures. This study paves the way for future applications of AI-driven models

390   in understanding antibody interactions with NCAAs. The success of our approach, which did not

391   require a starting structure, highlights the potential of computational methods to predict complex

392   biomolecular interactions and advance drug design and biomolecular research.

393   In summary, by combining experimental and AI-based structural data from RA-derived

394   monoclonal ACPAs, our studies provide first evidence that RA most likely initiates from a

395   repertoire of germline-encoded antibodies reactive to carbamylated proteins. In the context of RA

396   susceptibility, chronic antigen stimulation and antigen-driven affinity maturation, these antibodies

397   mature into a mixture of anti-CarP, ACPA and double reactive anti-CarP/ACPA, which are

398   markers that precede the clinical onset of RA.

399

400   **Methods**

401   **Cloning of monoclonal antibodies.** The HC and LC variable gene sequences of

402   monoclonal antibodies E4 (PDB: 5OD0_H, 50D0_L), T2H12 (GenBank: MH629725.1,

403   MH629693.1), T2E04 (GenBank: MH629730.1, MH629692.1), T3G09 (GenBank: MH629731.1,

404   MH629694.1) and T2D09 (GenBank: MH629717.1, MH629691.1), RA3 (GenBank: AJ430751.1,

405   AJ430766.1), F12-2 (GenBank: AJ430750.1, AJ430769.1), A2-2 (GenBank: AJ430749.1,

406   AJ430773.1), C3-4 (GenBank: AJ430740.1, AJ430759.10), G7-6 (GenBank: AJ430737.1,

407   AJ430756.1), A3-6 (GenBank: AJ430743.1, AJ430762.1), B3-6 (GenBank: AJ430752.1,

408   AJ430754.1) and C1-2 (GenBank: AJ430739.1, AJ430772.1) were synthesized through

409   Integrated DNA Technologies (IDT), and cloned into expression vectors containing human Igγ1,

410   Igκ or Igλ constant regions (kindly provided by Eric Meffre, Yale University School of Medicine,

411   New Haven, CT).

412   **Reversion of SHM sequences to germline.** The V(D)J germline sequences with the

413   lowest number of mismatch nucleotides compared to mutated sequences were obtained using

414   IMGT and IgBLAST, synthetized through IDT, and cloned into expression vectors containing

415   human Igγ1, Igκ or Igλ constant regions. In Gr1GL, the putative germline VH and VL CDR3s were

416   obtained from clade RA14 [47]. In Gr2GL, the putative CDR3s were determined by the most

417   conserved residues in antibodies C3-4, G7-6, C1-2, A3-6 and B3-6 in CDR H3

418   (ARGLSIYGTNDAFDV) and CDR L3 (AAWDDSLNGWV).

419   **Site-directed mutagenesis.** A single-point mutation from tryptophan H-47/48/49 to

420   methionine in antibodies E4, T3G09, T2D09, G7-6 and RA3 was performed using the Q5 Site-

421   Directed Mutagenesis Kit from New England BioLabs. Mutations were confirmed by Sanger

422   sequencing.

423   **Monoclonal antibody production.** Monoclonal antibodies were produced using the

424   ExpiCHO Expression System (Gibco) by co-transfecting plasmids encoding IgH and IgL

425   according to the manufacturer's instructions. Supernatants were collected on day 10 post-

426   transfection, and antibodies were purified using Protein A beads (Pierce).

427   **Preparation of Citrullinated and Carbamylated HEK-293T Cell Lysates.** HEK-293T

428   cells were transiently transfected with mock or plasmids encoding PAD2 and PAD4 using

429   Lipofectamine 2000 (Invitrogen). After 48 hours, the cells were incubated with 5 μM ionomycin

430   and 2 mM calcium in HBSS supplemented with 10mM HEPES for 2 hours at 37°C. Following

431   incubation, the cells were pelleted, washed twice with 1× PBS, resuspended in SDS sample

432   buffer, and boiled. For carbamylation, HEK-293T cells were resuspended in 1 M potassium

433   cyanate and sonicated in the presence of 1% Halt Protease Inhibitor Cocktail (Thermo Fisher).

434   After overnight incubation at 37°C, the sample was extensively dialyzed against 1× PBS.

435   Citrullination and carbamylation were confirmed by anti-modified citrulline (AMC) immunoblotting

436   [61].

437   **Immunoblotting.** Control, citrullinated and carbamylated cell lysates, as well as bovine

438   serum albumin (BSA) (ThermoFisher), acetylated BSA, alpha-1-antitrypsin (A1AT), and

439   carbamylated A1AT (Cayman Chemical) were analyzed by electrophoresis on 12% (lysates) or

440   10% (purified proteins) SDS-polyacrylamide gels, followed by protein transfer to nitrocellulose

441   membranes for immunoblotting. Except for monoclonal RA3M, which was used at 0.67 nM, all

442   RA-derived monoclonal antibodies were tested at a concentration of 6.7 nM. Images were

443   acquired using a ProteinSimple Fluorochem-M digital imager.

444   **Biomolecular complex prediction.** This section outlines the methods used for predicting

445   biomolecular complexes based on given sequence representations. All biomolecular structure

446   prediction methods investigated utilize a sequence representation of the wildtype or germline

447   antibody sequence, along with a Simplified Molecular Input Line Entry System (SMILES) string of

448   the citrullinated or carbamylated peptide. The peptide SMILES strings are generated using

449   PepSMI [62].

450   Wildtype heavy chain:

451   QVQLEESGPGLVRPSETLSLSCTVSGFPMSESYFWGWIRQSPGKGLEWLGSVIHTGTTYYRP

452   SLESRLTIAMDPSKNQVSLSLTSVTVADSAMYYCVRIRGGSSNWLDPWGPGIVVTASS

453   Wildtype light chain:

454 QSVWTQPPSVSAAPGQKVTISCSGDDSILRSAFVSWYQQVPGSAPKLVIFDDRQRPSGIPARF

455 SGSNSGTTATLDIAGLQRGDEADYYCAAWNGRLSAFVFGSGTKL

456 <u>Germline heavy chain:</u>

457 QVQLQESGPGLVKPSETLSLTCTVSGYSISSGYYWGWIRQPPGKGLEWIGSIYHSGSTYYNPS

458 LKSRVTISVDTSKNQFSLKLSSVTAADTAVYYCARIRGGSSNWFDPWGQGTLVTVSS

459 <u>Germline light chain:</u>

460 QSVLTQPPSVSAAPGQKVTISCSGSSSNIGNNYVSWYQQLPGTAPKLLIYDNNKRPSGIPDRFS

461 GSKSGTSATLGITGLQTGDEADYYCGTWDSSLSAFVFGTGTKV

462 <u>CII-Cit-13 peptide (GEEGK-Cit-GARG):</u>

463 NCC(=O)N<u>C@@</u>(CCC(=O)O)C(=O)N<u>C@@</u>(CCC(=O)O)C(=O)NCC(=O)N<u>C@@</u>(CCCCN)C(=O

464)N<u>C@@</u>(CCCNC(N)=O)C(=O)NCC(=O)N<u>C@@</u>(C)C(=O)N<u>C@@</u>(CCCNC(=N)N)C(=O)NCC(=O

465)O

466 <u>A1At-HCit peptide (EEAPLKLSKAVH-HCit-AVLTIDEKGTEAAGA):</u>

467 N<u>C@@</u>(CCC(=O)O)C(=O)N<u>C@@</u>(CCC(=O)O)C(=O)N<u>C@@</u>(C)C(=O)N1<u>C@@</u>(CCC1)C(=O)N

468 <u>C@@</u>(CC(C)C)C(=O)N<u>C@@</u>(CCCCN)C(=O)N<u>C@@</u>(CC(C)C)C(=O)N<u>C@@</u>(CO)C(=O)N<u>C@@</u>

469 (CCCCNC(N)=O)C(=O)N<u>C@@</u>(C)C(=O)N<u>C@@</u>(C(C)C)C(=O)N<u>C@@</u>(CC1=CN=C-

470 N1)C(=O)N<u>C@@</u>(CCCCN)C(=O)N<u>C@@</u>(C)C(=O)N<u>C@@</u>(C(C)C)C(=O)N<u>C@@</u>(CC(C)C)C(=O

471)N<u>C@@</u>(<u>C@</u>(O)C)C(=O)N<u>C@@</u>(<u>C@</u>(CC)C)C(=O)N<u>C@@</u>(CC(=O)O)C(=O)N<u>C@@</u>(CCC(=O)O

472)C(=O)N<u>C@@</u>(CCCCN)C(=O)NCC(=O)N<u>C@@</u>(<u>C@</u>(O)C)C(=O)N<u>C@@</u>(CCC(=O)O)C(=O)N<u>C</u>

473 <u>@@</u>(C)C(=O)N<u>C@@</u>(C)C(=O)NCC(=O)N<u>C@@</u>(C)C(=O)O

474   The reference crystal structure for validation tests is derived by truncating the 5OCX PDB

475 entry of E4 – CII-Cit-13 to only the variable domain of the antibody [38].

476   Predictions using various structure prediction methods were performed as follows: Chai-

477 1: Predictions are conducted with multiple sequence alignments (MSAs) enabled, no template,

478 no specific restraints, and default flags (3 trunk recycles and 200 diffusion timesteps) [40].

479 AlphaFold3: Predictions are made with default settings (10 recycles, 1 diffusion sample, and 1

480   seed) [41]. Boltz-1: Predictions are completed using MSA mode "mmseqs2_uniref_env," with 5

481   recycles, 200 sampling steps, and 1 diffusion sample [44]. RoseTTAFold All-Atom: Predictions are

482   executed with MSA search enabled [42]. Protenix: Predictions are performed with MSA search and

483   ESM features [43].

484   **Calculation of Root Mean Square Deviations (RMSD) for CDR Loops Compared to**

485   **Crystal Structures.** To assess the accuracy of the predicted complementarity determining region

486   (CDR) loops, we calculate the root mean square deviations (RMSD) between the predicted

487   structures and the corresponding crystal structure.

488   The RMSD for CDR loops and the framework region is calculated using the PyRosetta library [63].

489   The command to execute the RMSD calculation is as follows:

490

```
python rmsd.py --crystal [PATH_TO_CRYSTAL_PDB] --predicted
        [PATH_TO_PREDICTED_PDB]
```

492   In this command:

493   `--crystal` specifies the path to the crystal structure PDB file.

494   `--predicted` specifies the path to the predicted structure PDB file.

495   **Calculation of Global Biophysical Properties.** Global biophysical properties of the

496   antibodies E4 and E4GL, including net charge, paratope charge, paratope solvent accessible

497   surface area (SASA), and VH/VL packing angles and distances, are calculated using the Rosetta

498   AntibodyFeatures reporter [64].

499   To execute the calculation, use the following command:

500

```
python AntibodyFeatures.py [PATH_TO_UNBOUND_PDB]
```

501   **Calculation of Per-CDR Biophysical Properties.** Per-CDR biophysical properties of E4

502   and E4GL are derived by querying the SQL database generated from the AntibodyFeatures

503   reporter.

504   To perform this analysis, run the following command:

505

```
python db3_check.py --db3 [PATH_TO_DB3]
```

506 **Calculation of Binding Affinity for Antibody-Peptide Complexes.** Binding affinity

507 between antibody-peptide complexes is predicted using PRODIGY-LIG, a structure-based

508 method designed for the prediction of binding affinity in protein-small ligand complexes [65].

509 The binding affinity is calculated using the following command:

510 `prodigy_lig -c A,B,C C:LIG -i [PATH_TO_BOUND_PDB]`

511 In this command:

512 `--c A,B,C` specifies the chains of the antibody involved in the interaction

513 `C:LIG` indicates that the ligand (peptide) is associated with chain C

514 `--i [PATH_TO_BOUND_PDB]` is the path to the PDB file containing the antibody-peptide

515 complex. The default distance cutoff for atomic contacts is set to 10.5 Å, which has been

516 established as optimal for these calculations.

517 **Identification of Intermolecular Interactions.** Intermolecular interactions between the

518 antibody and the peptide are visualized using PyMol. Specifically, the citrulline (or homocitrulline)

519 within the peptide is selected, and the following command identifies any polar contacts between

520 the residue and the heavy or light chain of the antibody:

521 `actions > find > polar contacts > to any atoms`

522

523

References

1. Rantapaa Dahlqvist, S. & Andrade, F. Individuals at risk of seropositive rheumatoid arthritis: the evolving story. *J Intern Med* **286**, 627-643 (2019).

2. Trejo-Zambrano, M.I., Gomez-Banuelos, E. & Andrade, F. Redox-Mediated Carbamylation As a Hapten Model Applied to the Origin of Antibodies to Modified Proteins in Rheumatoid Arthritis. *Antioxid Redox Signal* (2021).

3. Schellekens, G.A., de Jong, B.A., van den Hoogen, F.H., van de Putte, L.B. & van Venrooij, W.J. Citrulline is an essential constituent of antigenic determinants recognized by rheumatoid arthritis-specific autoantibodies. *J. Clin. Invest* **101**, 273-281 (1998).

4. Shi, J.*, et al.* Autoantibodies recognizing carbamylated proteins are present in sera of patients with rheumatoid arthritis and predict joint damage. *Proc. Natl. Acad. Sci. U. S. A* **108**, 17372-17377 (2011).

5. Rantapaa-Dahlqvist, S.*, et al.* Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. *Arthritis Rheum* **48**, 2741-2749 (2003).

6. Brink, M.*, et al.* Anti-carbamylated protein antibodies in the pre-symptomatic phase of rheumatoid arthritis, their relationship with multiple anti-citrulline peptide antibodies and association with radiological damage. *Arthritis Res Ther* **17**, 25 (2015).

7. Yu, K.R. & Proost, P. Insights into peptidylarginine deiminase expression and citrullination pathways. *Trends Cell Biol* **32**, 746-761 (2022).

545 8.   Darrah, E. & Andrade, F. Rheumatoid arthritis and citrullination. *Curr. Opin. Rheumatol*
546   **30**, 72-78 (2018).

547 9.   Stark, G.R., Stein, W.H. & Moore, S. Reactions of Cyanate Present in Aqueous Urea with
548   Amino Acids and Proteins. *Journal of Biological Chemistry* **235**, 3177-3181 (1960).

549 10.   Stark, G.R. On the Reversible Reaction of Cyanate with Sulfhydryl Groups and the
550   Determination of Nh2-Terminal Cysteine and Cystine in Proteins. *J Biol Chem* **239**, 1411-
551   1414 (1964).

552 11.   Stark, G.R. Reactions of cyanate with functional groups of proteins. 3. Reactions with
553   amino and carboxyl groups. *Biochemistry* **4**, 1030-1036 (1965).

554 12.   Delanghe, S., Delanghe, J.R., Speeckaert, R., Van Biesen, W. & Speeckaert, M.M.
555   Mechanisms and consequences of carbamoylation. *Nat Rev Nephrol* **13**, 580-593 (2017).

556 13.   Lee, C.Y.*, et al.* Mining the Human Tissue Proteome for Protein Citrullination. *Mol Cell*
557   *Proteomics* **17**, 1378-1391 (2018).

558 14.   Kampstra, A.S.B.*, et al.* Different classes of anti-modified protein antibodies are induced
559   on exposure to antigens expressing only one type of modification. *Ann Rheum Dis* **78**,
560   908-916 (2019).

561 15.   Yamada, H.*, et al.* Cutting Edge: B Cells Expressing Cyclic Citrullinated Peptide-Specific
562   Antigen Receptor Are Tolerized in Normal Conditions. *J Immunol* **201**, 3492-3496 (2018).

563   16.   Dwivedi, N*., et al.* B Cell Tolerance to Deiminated Histones in BALB/c, C57BL/6, and

564   Autoimmune-Prone Mouse Strains. *Front Immunol* **8**, 362 (2017).

565   17.   McElwee, M.K., Dileepan, T., Mahmud, S.A. & Jenkins, M.K. The CD4+ T cell repertoire

566   specific for citrullinated peptides shows evidence of immune tolerance. *J Exp Med*

567   **220**(2023).

568   18.   Hill, J.A*., et al.* Arthritis induced by posttranslationally modified (citrullinated) fibrinogen in

569   DR4-IE transgenic mice. *J. Exp. Med* **205**, 967-979 (2008).

570   19.   Rubin, B. & Sonderstrup, G. Citrullination of self-proteins and autoimmunity. *Scandinavian

571   Journal of Immunology* **60**, 112-120 (2004).

572   20.   Mydel, P*., et al.* Carbamylation-dependent activation of T cells: a novel mechanism in the

573   pathogenesis of autoimmune arthritis. *J. Immunol* **184**, 6882-6890 (2010).

574   21.   Ho, P.P*., et al.* Autoimmunity against fibrinogen mediates inflammatory arthritis in mice. *J.

575   Immunol* **184**, 379-390 (2010).

576   22.   Cantaert, T., Teitsma, C., Tak, P.P. & Baeten, D. Presence and role of anti-citrullinated

577   protein antibodies in experimental arthritis models. *Arthritis Rheum* **65**, 939-948 (2013).

578   23.   Carracedo, J*., et al.* Protein Carbamylation: A Marker Reflecting Increased Age-Related

579   Cell Oxidation. *Int J Mol Sci* **19**(2018).

580   24.   Gorisse, L*., et al.* Protein carbamylation is a hallmark of aging. *Proc Natl Acad Sci U S A*
581   **113**, 1191-1196 (2016).

582   25.   Wang, Z*., et al.* Protein carbamylation links inflammation, smoking, uremia and
583   atherogenesis. *Nat. Med* **13**, 1176-1184 (2007).

584   26.   Delporte, C*., et al.* Myeloperoxidase-catalyzed oxidation of cyanide to cyanate: A potential
585   carbamylation route involved in the formation of atherosclerotic plaques? *J Biol Chem* **293**,
586   6374-6386 (2018).

587   27.   Gillery, P. & Jaisson, S. Post-translational modification derived products (PTMDPs): toxins
588   in chronic diseases? *Clin Chem Lab Med* **52**, 33-38 (2014).

589   28.   Gillery, P. & Jaisson, S. Usefulness of non-enzymatic post-translational modification
590   derived products (PTMDPs) as biomarkers of chronic diseases. *J Proteomics* **92**, 228-238
591   (2013).

592   29.   Turunen, S., Koivula, M.K., Risteli, L. & Risteli, J. Anticitrulline antibodies can be caused
593   by homocitrulline-containing proteins in rabbits. *Arthritis Rheum* **62**, 3345-3352 (2010).

594   30.   Wells, H.G. *The chemical aspects of immunity*, (Chemical Catalog Co., New York, 1929).

595   31.   Steinbrecher, U.P., Fisher, M., Witztum, J.L. & Curtiss, L.K. Immunogenicity of
596   homologous low density lipoprotein after methylation, ethylation, acetylation, or
597   carbamylation: generation of antibodies specific for derivatized lysine. *J Lipid Res* **25**,
598   1109-1116 (1984).

599   32.   Dekkers, J.S.*, et al.* Breach of autoreactive B cell tolerance by post-translationally modified

600   proteins. *Ann Rheum Dis* **76**, 1449-1457 (2017).

601   33.   Verheul, M.K., Vierboom, M.P.M., t Hart, B.A., Toes, R.E.M. & Trouw, L.A. Anti-

602   carbamylated protein antibodies precede disease onset in monkeys with collagen-induced

603   arthritis. *Arthritis Res Ther* **19**, 246 (2017).

604   34.   Sahlstrom, P.*, et al.* Different Hierarchies of Anti-Modified Protein Autoantibody

605   Reactivities in Rheumatoid Arthritis. *Arthritis Rheumatol* **72**, 1643-1657 (2020).

606   35.   Figueiredo, C.P.*, et al.* Antimodified protein antibody response pattern influences the risk

607   for disease relapse in patients with rheumatoid arthritis tapering disease modifying

608   antirheumatic drugs. *Ann Rheum Dis* **76**, 399-407 (2017).

609   36.   Juarez, M.*, et al.* Identification of novel antiacetylated vimentin antibodies in patients with

610   early inflammatory arthritis. *Ann Rheum Dis* **75**, 1099-1107 (2016).

611   37.   van de Stadt, L.A.*, et al.* Monoclonal anti-citrullinated protein antibodies selected on

612   citrullinated fibrinogen have distinct targets with different cross-reactivity patterns.

613   *Rheumatology. (Oxford)* **52**, 631-635 (2013).

614   38.   Ge, C.*, et al.* Structural Basis of Cross-Reactivity of Anti-Citrullinated Protein Antibodies.

615   *Arthritis Rheumatol* **71**, 210-221 (2019).

616   39.   He, Y.*, et al.* A subset of antibodies targeting citrullinated proteins confers protection from

617   rheumatoid arthritis. *Nat Commun* **14**, 691 (2023).

618   40.   Boitreaud, J.*, et al.* Chai-1: Decoding the molecular interactions of life. *bioRxiv [preprint]*

619   **doi.org/10.1101/2024.10.10.615955** (2024).

620   41.   Abramson, J.*, et al.* Accurate structure prediction of biomolecular interactions with

621   AlphaFold 3. *Nature* **630**, 493-500 (2024).

622   42.   Krishna, R.*, et al.* Generalized biomolecular modeling and design with RoseTTAFold All-

623   Atom. *Science* **384**, eadl2528 (2024).

624   43.   Chen, X.*, et al.* Protenix - Advancing Structure Prediction Through a Comprehensive

625   AlphaFold3 Reproduction. *bioRxiv [preprint]* **doi: 10.1101/2025.01.08.631967** (2025).

626   44.   Wohlwend, J.*, et al.* Boltz-1 Democratizing Biomolecular Interaction Modeling. *bioRxiv*

627   *[preprint]* **DOI: 10.1101/2024.11.19.624167** (2024).

628   45.   Verheul, M.K.*, et al.* Identification of carbamylated alpha 1 anti-trypsin (A1AT) as an

629   antigenic target of anti-CarP antibodies in patients with rheumatoid arthritis. *J Autoimmun*

630   **80**, 77-84 (2017).

631   46.   Shehata, L.*, et al.* Affinity Maturation Enhances Antibody Specificity but Compromises

632   Conformational Stability. *Cell Rep* **28**, 3300-3308 e3304 (2019).

633   47.   Titcombe, P.J.*, et al.* Pathogenic Citrulline-Multispecific B Cell Receptor Clades in

634   Rheumatoid Arthritis. *Arthritis Rheumatol* **70**, 1933-1945 (2018).

635 48. Raats, J.M., Wijnen, E.M., Pruijn, G.J., van den Hoogen, F.H. & van Venrooij, W.J.
636   Recombinant human monoclonal autoantibodies specific for citrulline-containing peptides
637   from phage display libraries derived from patients with rheumatoid arthritis. *J Rheumatol*
638   **30**, 1696-1711 (2003).

639 49. Kissel, T*., et al.* Antibodies and B cells recognising citrullinated proteins display a broad
640   cross-reactivity towards other post-translational modifications. *Ann Rheum Dis* **79**, 472-
641   480 (2020).

642 50. Kongpachith, S*., et al.* Affinity Maturation of the Anti-Citrullinated Protein Antibody
643   Paratope Drives Epitope Spreading and Polyreactivity in Rheumatoid Arthritis. *Arthritis*
644   *Rheumatol* **71**, 507-517 (2019).

645 51. Steen, J*., et al.* Recognition of Amino Acid Motifs, Rather Than Specific Proteins, by
646   Human Plasma Cell-Derived Monoclonal Antibodies to Posttranslationally Modified
647   Proteins in Rheumatoid Arthritis. *Arthritis Rheumatol* **71**, 196-209 (2019).

648 52. Shi, J*., et al.* Affinity maturation shapes the function of agonistic antibodies to
649   peptidylarginine deiminase type 4 in rheumatoid arthritis. *Ann Rheum Dis* **77**, 141-148
650   (2018).

651 53. Gomez-Banuelos, E*., et al.* Affinity maturation generates pathogenic antibodies with dual
652   reactivity to DNase1L3 and dsDNA in systemic lupus erythematosus. *Nat Commun* **14**,
653   1388 (2023).

654   54.   Kummu, O., *et al.* Carbamyl adducts on low-density lipoprotein induce IgG response in
655   LDLR-/- mice and bind plasma autoantibodies in humans under enhanced carbamylation.
656   *Antioxid Redox Signal* **19**, 1047-1062 (2013).

657   55.   Ospelt, C., *et al.* Carbamylation of vimentin is inducible by smoking and represents an
658   independent autoantigen in rheumatoid arthritis. *Ann Rheum Dis* **76**, 1176-1183 (2017).

659   56.   Wang, Z., *et al.* Eosinophil Peroxidase Catalyzed Protein Carbamylation Participates in
660   Asthma. *J Biol Chem* **291**, 22118-22135 (2016).

661   57.   Bright, R., *et al.* Gingival tissue, an extrasynovial source of malondialdehyde-acetaldehyde
662   adducts, citrullinated and carbamylated proteins. *J Periodontal Res* **53**, 139-143 (2018).

663   58.   Lee, Y.H., *et al.* Localisation of citrullinated and carbamylated proteins in inflamed gingival
664   tissues from rheumatoid arthritis patients. *Clin Oral Investig* **25**, 1441-1450 (2021).

665   59.   Gomez-Banuelos, E., Konig, M.F. & Andrade, F. Microbial pathways to subvert host
666   immunity generate citrullinated neoantigens targeted in rheumatoid arthritis. *Curr Opin*
667   *Struct Biol* **75**, 102423 (2022).

668   60.   Konig, M.F., *et al.* Aggregatibacter actinomycetemcomitans-induced hypercitrullination
669   links periodontal infection to autoimmunity in rheumatoid arthritis. *Sci Transl. Med* **8**,
670   369ra176 (2016).

61. Senshu, T., Sato, T., Inoue, T., Akiyama, K. & Asaga, H. Detection of citrulline residues in deiminated proteins on polyvinylidene difluoride membrane. *Anal. Biochem* **203**, 94-100 (1992).

62. PepSMI. *https://www.novoprolabs.com/tools/convert-peptide-to-smiles-string*.

63. Chaudhury, S., Lyskov, S. & Gray, J.J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689-691 (2010).

64. Adolf-Bryfogle, J*., et al.* RosettaAntibodyDesign (RAbD): A general framework for computational antibody design. *PLoS Comput Biol* **14**, e1006112 (2018).

65. Vangone, A*., et al.* Large-scale prediction of binding affinity in protein-small ligand complexes: the PRODIGY-LIG web server. *Bioinformatics* **35**, 1585-1587 (2019).

66. Jurrus, E*., et al.* Improvements to the APBS biomolecular solvation software suite. *Protein Sci* **27**, 112-128 (2018).

67. Madeira, F*., et al.* The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Research* **52**, W521-W525 (2024).

693

694   **Author contributions.**

695   Conception: FA. Designing research studies: MES, MC, JJG, FA. Conducting experiments: MES,

696   MC, DF, JJG, FA. Data acquisition and analyses: MES, MC, DF, JJG, FA. Interpretation of data:

697   MES, MC, DF, JJG, FA. Writing—original draft: MES, MC, JJG, FA. Writing—review and editing:

698   MES, MC, DF, JJG, FA. All authors reviewed, edited, and approved the manuscript.

699

700   **Competing interests:** None.

701

702   **Data and materials availability:** All data are available in the main text or the supplementary

703   materials.

704

705

Figures
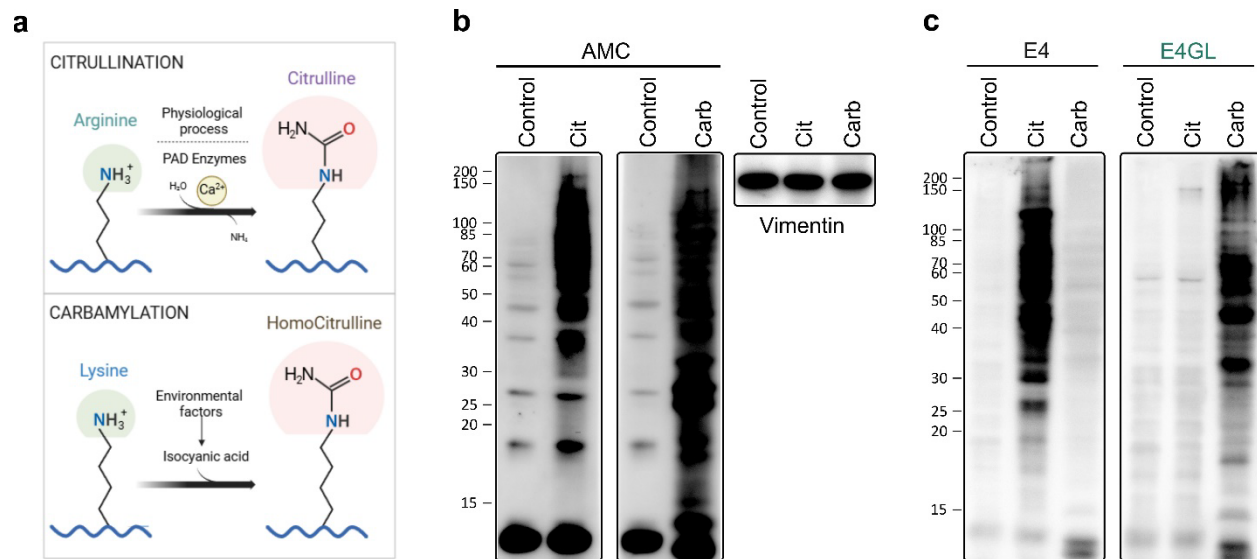
Fig. 1. Monoclonal ACPA E4 changes reactivity to carbamylated proteins when reverted to its germline form. (**a**) Schematic view of the chemical structure of peptidyl-arginine and peptidyl-lysine, and their conversions to citrulline and homocitrulline residues (upper and lower panel, respectively). The amino acid side chain group that is modified is circled in green, and the resulting modification in pink. Citrulline and homocitrulline both contain a ureido group in the modified side chain (NH_2CONH-). Homocitrulline is one methylene group longer than citrulline. PAD: Peptidylarginine deiminase. The figure was created with BioRender.com. (**b**) Mock or PAD2/PAD4 co-expressing 293T cells were stimulated with ionomycin to generate control and hypercitrullinated (Cit) cells, respectively. 293T cells were incubated with 1M potassium cyanate to induce carbamylation (Carb). Cellular citrullination and carbamylation were confirmed by AMC immunoblotting. Detection of vimentin was used as loading control (right panel). All immunoblotting assays used the same cell lysate batch and protein amount. (**c**) Lysates from control, Cit and Carb cells were tested by immunoblotting using 6.7 nM of antibody E4 or germline reverted E4 (E4GL). Data in **b** and **c** is representative from 2 and 4 independent experiments, respectively.
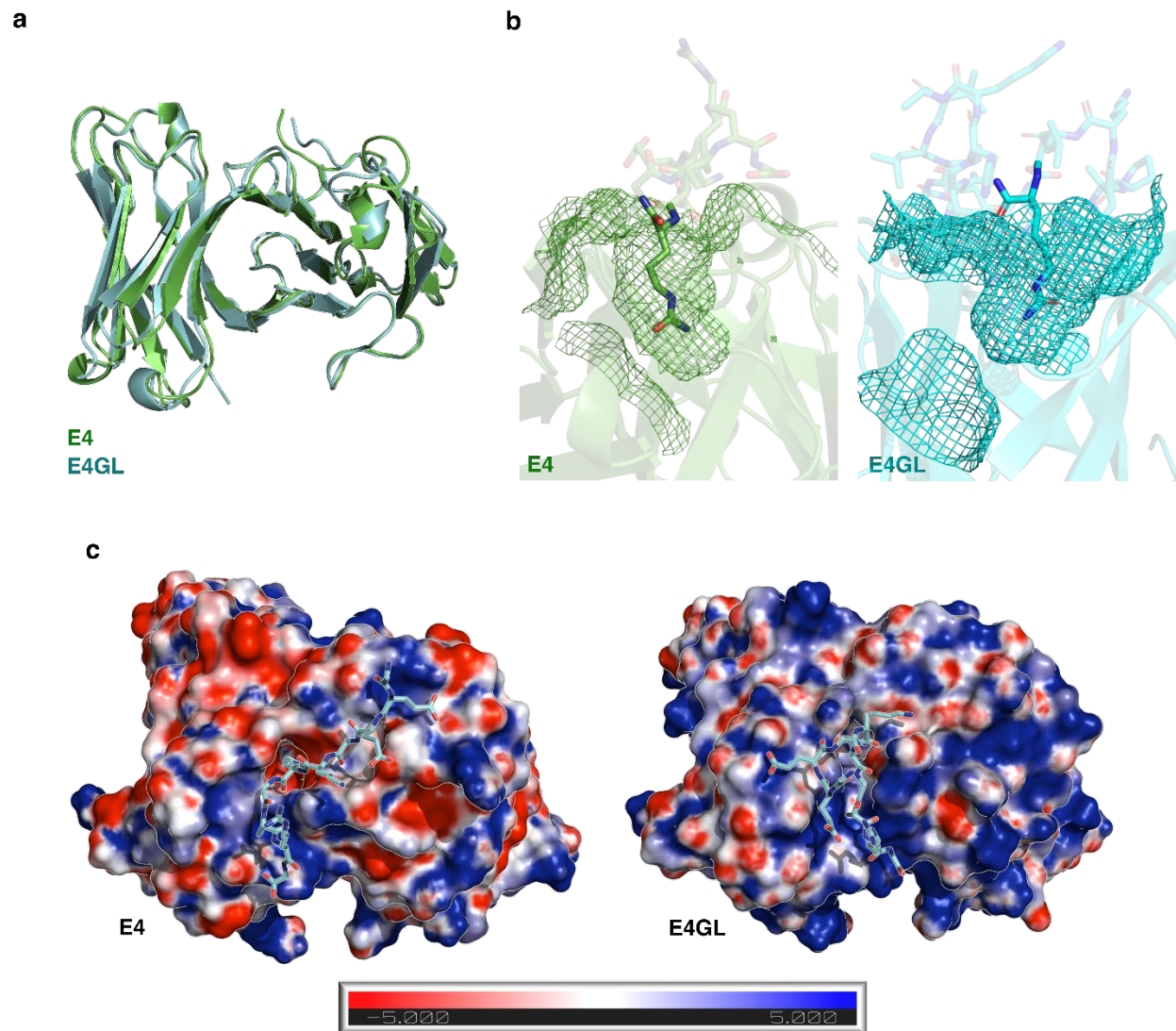
723

Fig. 2. Biophysical characterization of deduced structures of E4 and E4GL. (**a**) Comparison of E4 and E4GL variable domains predicted by Chai-1. (**b**) A mesh representation of the E4 paratope with CII-Cit-13 and E4GL with A1AT-HCit reveals the variation in each antibody's binding site architecture. (**c**) Charge-based surface models of E4 and E4GL. Protonation states calculated at physiological pH (7.4) with pdb2pqr [66]; electrostatics calculated with APBS [66]. The predicted CII-Cit-13 peptide is superimposed on the surface representation of each antibody to visualize approximate peptide placement relative to the charged patches.
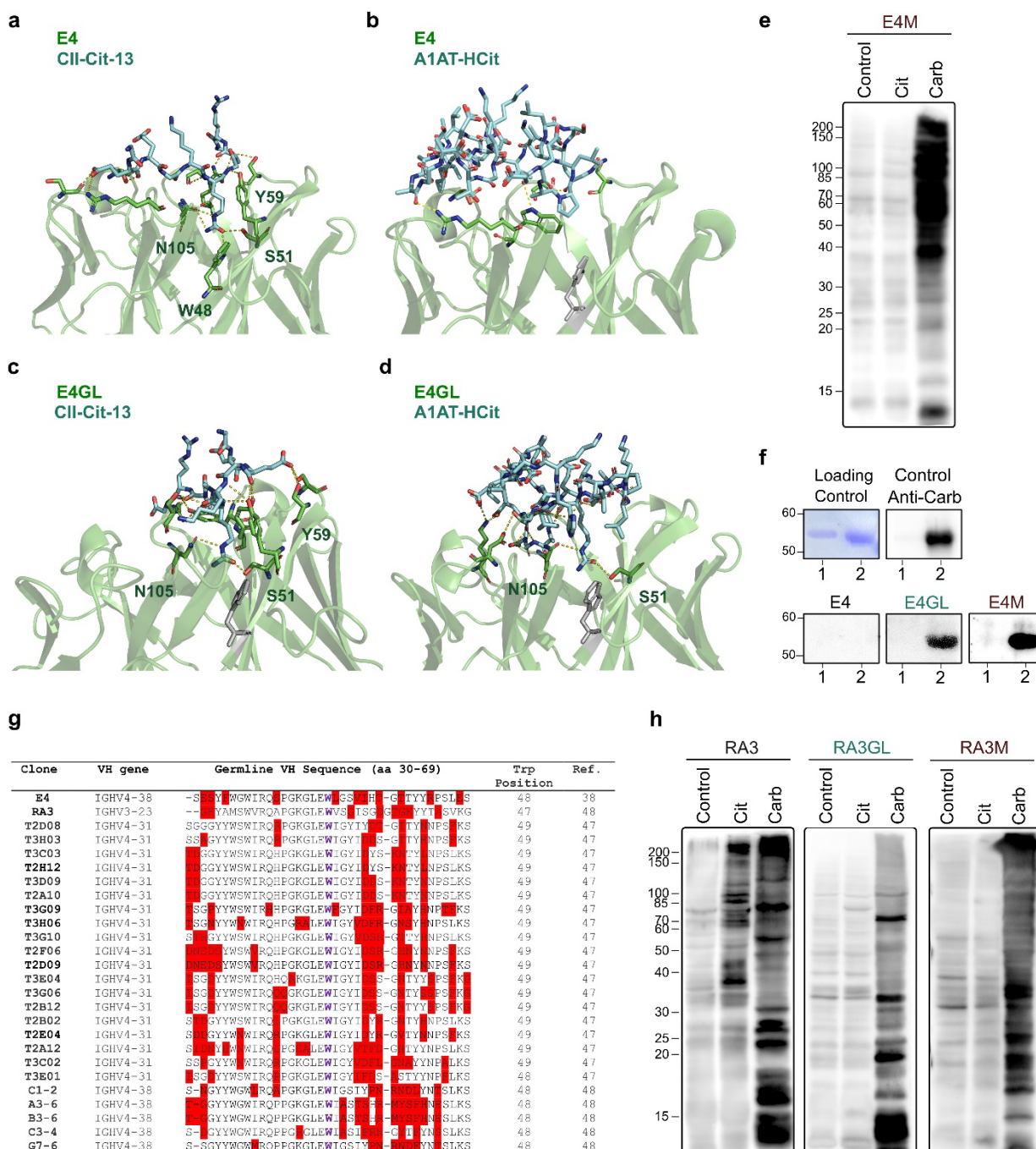
731

732

Fig. 3. Heavy-chain tryptophan 48 (H-W48) dictates E4 binding to citrullinated or carbamylated proteins. Deduced structures from Chai-1 of (**a**) E4 bound to CII-Cit-13, (**b**) E4 bound to A1AT-HCit, (**c**) E4GL bound to CII-Cit-13, and (**d**) E4GL bound to A1AT-HCit. All antibody paratope residues (green) that interact with the bound peptide (cyan) are shown in stick representation. For clarity, only the antibody residues that interact directly with citrulline or

738 homocitrulline are annotated. (**e**) Lysates from control, hypercitrullinated (Cit) and carbamylated

739 (Carb) cells were tested by immunoblotting using antibody E4 in which H-W48 was changed to

740 methionine (E4M). (**f**) Control (lane 1) and carbamylated (Carb, lane 2) α1-antitrypsin were

741 detected by Coomassie staining (loading control) or by immunoblotting with a commercial

742 antibody to anti-carbamyl-lysine (Cell Biolabs Inc., positive control) and antibodies E4, E4GL and

743 E4M. (**g**) Sequence alignment of heavy chain (IgH) amino acid residues 33-66 from published RA

744 patient-derived monoclonal ACPAs. Missense mutations are marked in red and conserved H-

745 W47/48/49 in purple. Sequences were aligned using the multiple sequence alignment tool Clustal

746 Omega from EMBL-EBI[67]. Except for monoclonal RA3M, which was used at 0.67 nM, all RA-

747 derived monoclonal antibodies were tested at a concentration of 6.7 nM. Data in **e**, **f** and **h** is

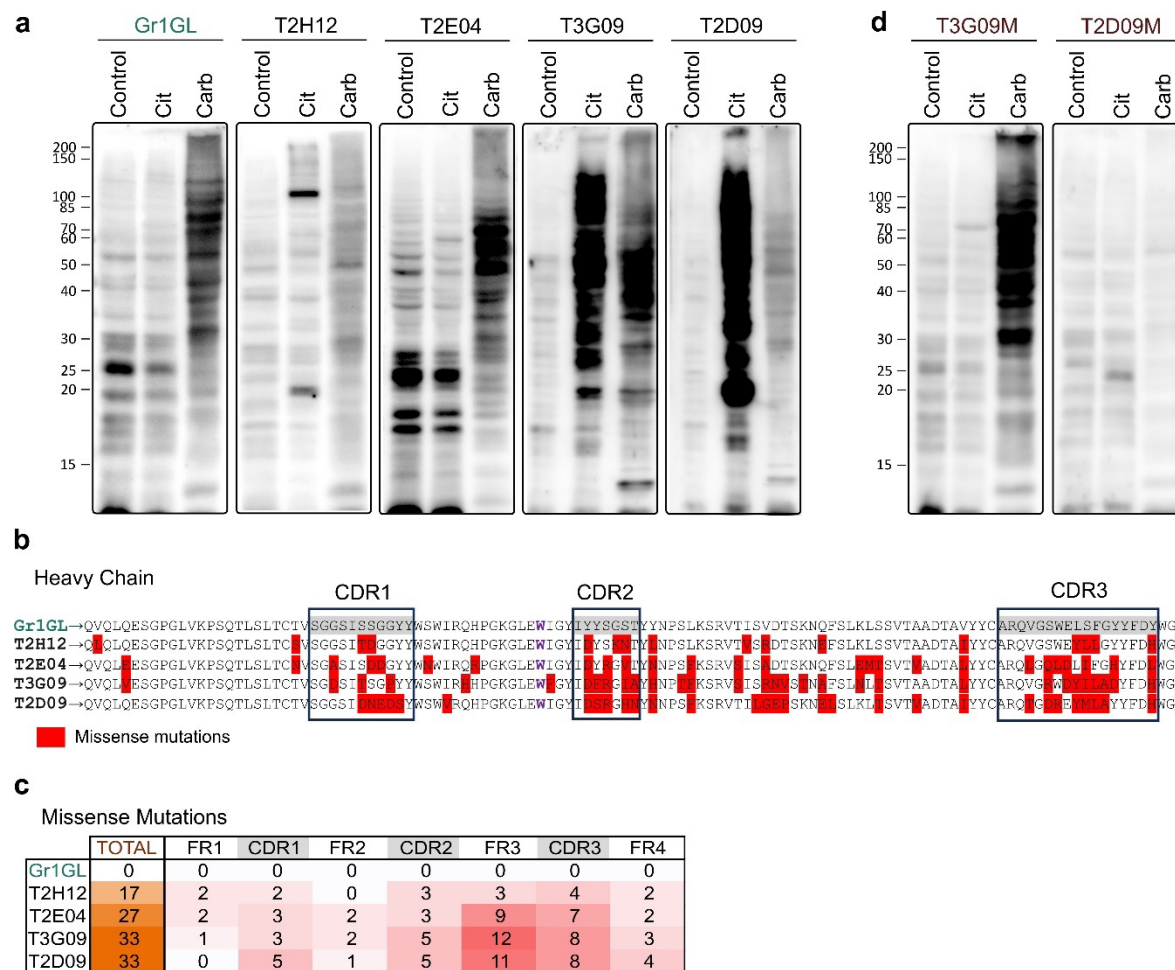748 representative from 3 independent experiments.

Fig. 4. ACPAs, anti-CarP and double reactive anti-CarP/ACPAs comprise antibodies with distinct number and pattern of somatic hypermutations derived from single anti-CarP germline clones. (**a**) Lysates from control, hypercitrullinated (Cit) and carbamylated (Carb) cells were tested by immunoblotting using antibody Gr1GL, T2H12, T2E04, T3G09 or T2D09. (**b, c**) VH amino acid sequence alignment (**b**) and number of missense mutations (**c**) of antibodies shown in **a**. In **b**, missense mutations are marked in red and H-W49 in purple. FR, framework region. CDR, complementarity-determining region. (**d**) Lysates from control, Cit and Carb cells were tested by immunoblotting using antibody T3G09M or T2D09M. All RA-derived monoclonal antibodies were tested at a concentration of 6.7 nM. Data in **a** and **d** is representative from 3 independent experiments.
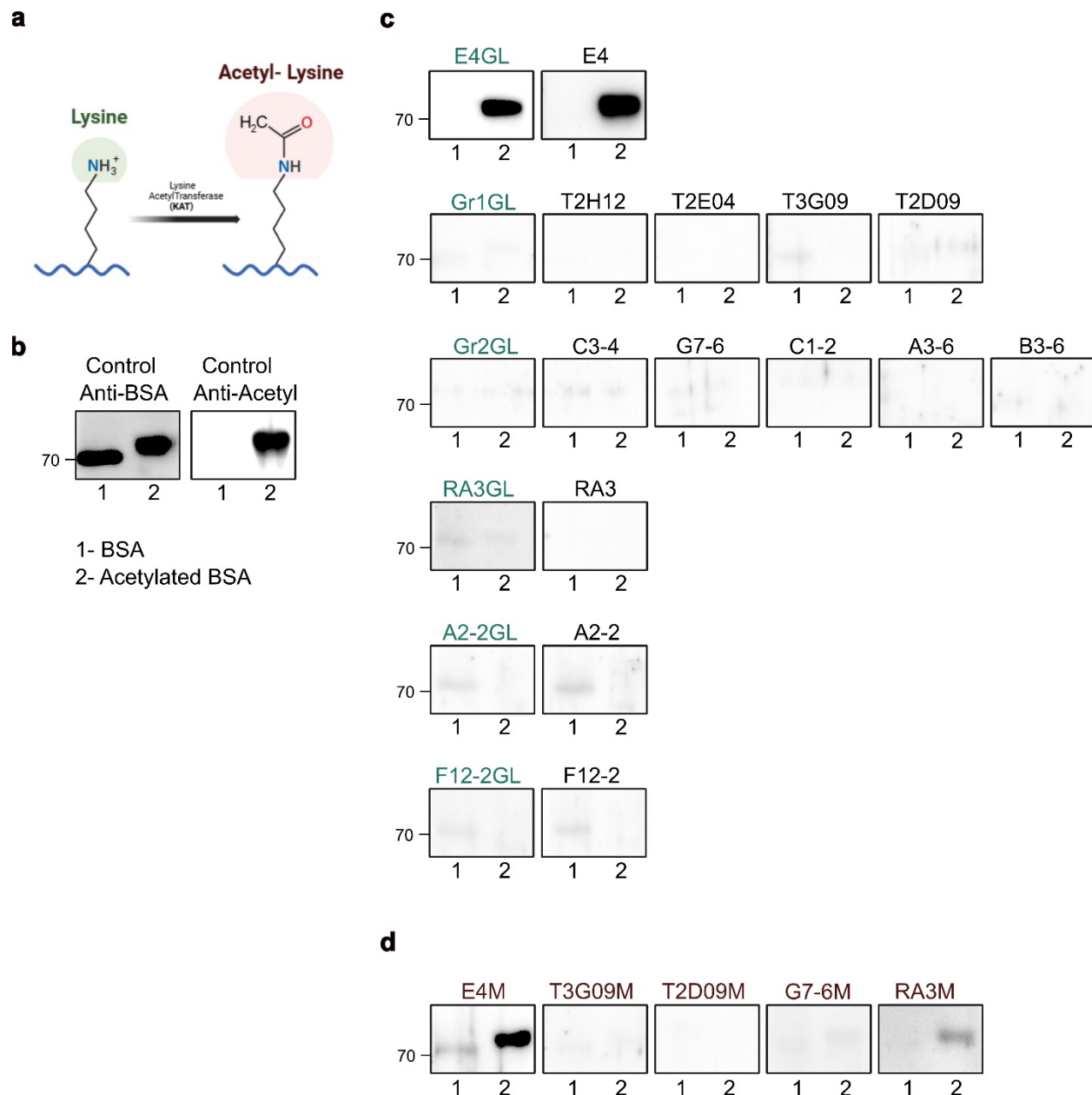
Fig. 5. Protein acetylation is a cross-reactive target for certain ACPAs. (**a**) Schematic representation of the chemical conversion of peptidyl-lysine to an acetyl-lysine residue. The amino acid side chain group that is modified is circled in green, and the resulting modification in pink. Acetyl-lysine differs from homocitrulline by having a methylene group in place of the terminal amine found in the ureido group of homocitrulline. The figure was created with BioRender.com. (**b**) Bovine serum albumin (BSA) (lane 1) and acetylated BSA (lane 2) were detected using an in-

768   house made anti-BSA mouse polyclonal antibody (loading control) or a commercial anti-

769   acetylated-lysine antibody (Millipore Sigma) as the positive control. (**c, d**) Control and acetylated

770   BSA were analyzed by immunoblotting using antibodies E4GL, E4, Gr1GL, T2H12, T2E04,

771   T3G09, T2D09, Gr2GL, C3-4, G7-6, C1-2, A3-6, B3-6, RA3GL, RA3, A2-2GL, A2-2, F12-2GL and

772   F12-2 **(c)** and H-M47/48/49 mutants **(d)**. Except for monoclonal RA3M, which was used at 0.67

773   nM, all RA-derived monoclonal antibodies were tested at a concentration of 6.7 nM. Data in **c** and

774   **d** are representative of 2 independent experiments.

775

776