

COMMUNICATION

A QM/MM Study on the Origin of Retro-Aldolase Activity of a Catalytic Antibody.

Daria De Raffele,^a Sergio Martí^a and Vicent Moliner^{*a}

Received 00th February 20xx,
Accepted 00th February 20xx

DOI: 10.1039/x0xx00000x

The retro-aldolase mechanism of methodol catalysed by the catalytic antibody 33F12 is described based on the exploration of the free energy landscape obtained with QM/MM methods. The amino acids involved in the reaction have been identified, as well as their specific role played in the active site and in the flexibility of the loops. Finally, the comparison with a *de novo* enzyme RA95.5-8F provides a deeper understanding of catalytic differences between such different protein scaffolds.

In 1986 Richard A. Lerner and Alfonso Tramontano began their article Catalytic Antibodies (CAs) making the question: "Can antibodies be made to serve as enzymes?".¹ After only five years from the publication of this article, a substantial number of reactions were successfully catalysed by antibodies: carbonate hydrolysis, ester hydrolysis, Claisen rearrangement, transesterification or Diels-Alder reactions, just to mention some of them.² The capability to produce such CAs comes from the proper design of a stable molecule, which must resemble the structure of the transition state (TS) of the chemical reaction (transition state analogue, TSA). Then, the TSA is used as a hapten to challenge an immune system, which replies generating a germline antibody that will bind it and, hopefully, it would also present affinity for the real TS. An affinity maturation process is followed, where catalytic groups can be introduced directly by chemical modification or by site-directed mutagenesis or genetic selection, to lead to a high affinity monoclonal matured antibody with catalytic properties.³ Nevertheless, two serious drawbacks appear: first, because the TSA does not correspond to the real protein–substrate TS, the specific pattern of interactions established with the CA does not exactly match to the one found in the TS; and second, it has been claimed that the high binding affinity developed by the antibody during its maturation against the TSA molecule provides an

excessive rigidity to the Michaelis complex, which translates into a loss of plasticity compared to natural enzymes. These limitations can be in the origin of the usually low efficiency of the CAs, which combined with the poor knowledge of the molecular mechanism of the chemical reactions inside the protein^{4–7}, hamper the progress in the application of CAs as catalysts. Anyway, due to the ability to yield antibodies tailored for any specific reaction, they have been widely studied for different purposes ranging from therapeutic or diagnostic agents to the creation of novel materials.^{8,9} Originally, the existence of enzymes that catalyse the aldol condensation was well known, despite with a limited number of substrates. In 1995 Barbas III and co-workers published a study which goal was to generate CAs to catalyse the aldol reaction by means of the mechanism used by natural class I aldolase enzymes, but with the availability of using different substrates.¹⁰ Out of 20 antibodies screened, two of them, 38C2 and 33F12, presented not only aldolase catalytic activity but they were able to catalyse over 100 different reactions, including retro-aldol, and enamine / iminium-based transformations of a very wide range of substrates and stereochemical specificities.^{10,11} In order to better

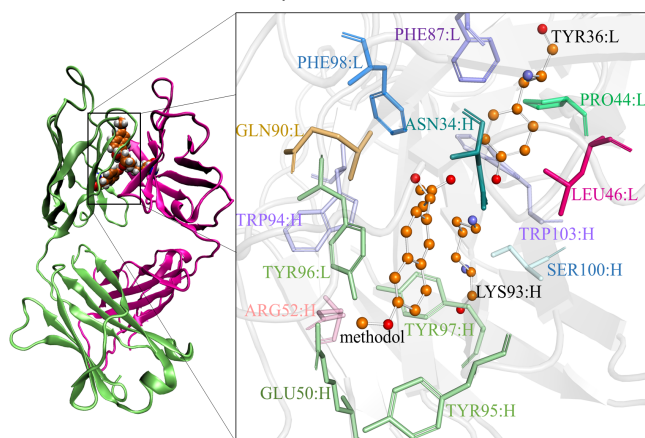


Fig. 1 Representation of the X-ray structure of L (magenta) and H (green) chains of 33F12 and detail of the active site in VDW representation refers to residues that take part in the reaction.

^a Departament de Química Física i Analítica; Universitat Jaume I. 12071 Castellón (Spain).

† Footnotes relating to the title and/or authors should appear here.

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

understand the structural features of the active site of these two sequence-related variants, both CAs were cloned and their genes sequenced.¹¹ The results revealed that the two CAs were somatic variants of a single VDJ rearrangement and differed by just nine amino acids each in the two variable regions V_L and V_H .¹¹ The analysis of the X-ray crystal structure of 33F12 showed that the entry is a narrow-elongated cleft, and the binding pocket is more than 11 Å deep. The reactive Lys93:H is located at the bottom of the pocket, within a hydrophobic environment. Apparently, the absence of charged functional groups or hydrogen bonds in the surroundings of Lys93:H promotes its deprotonated state making it a strong nucleophile.¹² In the present work, the retro-aldol reaction of the non-natural substrate 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone (methodol) catalysed by CA 33F12 (see Fig. 1), whose full X-ray structure is available,¹³ is carried out by means of Quantum Mechanics/Molecular Mechanics (QM/MM) methods. It is important to point out that despite there are no evidences on the ability of 33F12 to discriminate between enantiomeric substrates, either 38C2 and the *de novo* enzymes explored in our previous studies^{14,15} show (R) stereospecificity. The choice of the retro-aldol conversion lies in its broad used in biochemical syntheses because of the carbon-carbon bond breaking availability and the fact that there is no natural enzyme known that catalyses this reaction. In addition, many efforts have been directed to designing new enzymes with retro-aldol activity, which will allow carrying out comparisons between CA and *de novo* designs. In this regard, Baker et al. in 2008 designed a total of 72 enzymes, 32 of which showed retro-aldol catalytic activity against methodol¹⁶. Indeed, the best resulting enzyme, RA95.5-8F, showed rate enhancements of up to 10^9 over the uncatalyzed reaction.¹⁷ This discovery led to several studies to detect its chemical, thermodynamic and kinetic characteristics.^{14,18,19} Notwithstanding, practically all *de novo* enzymes are still far from resembling the catalytic skills of their natural counterparts. The study of CAs, based on a completely different protein scaffold, can also be useful to improve not only their own catalytic efficiency but can provide information to improve *de novo* enzymes. Our results will allow identifying the overall molecular mechanism, and investigating the role played by the

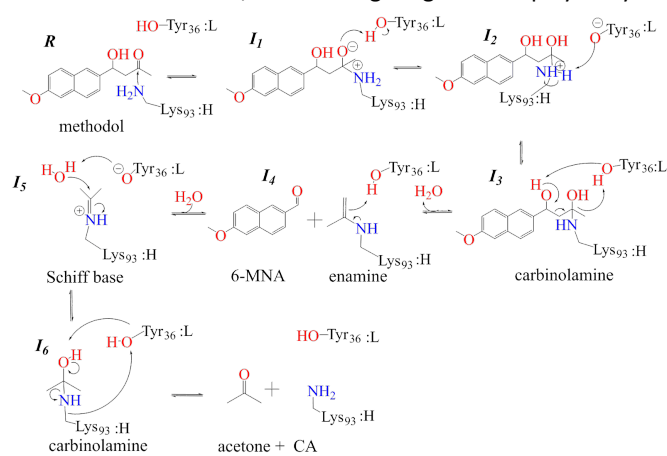


Fig. 2 Schematic representation of the proposed reaction mechanism of the retro-aldol reaction on methodol, catalyzed by the catalytic antibody 33F12.

key amino acids that participate in the reaction. Moreover, a deep comparative analysis with the results previously obtained for the same reaction catalysed by the *de novo* RA95.5-8F enzyme (Fig. S3) will provide valuable information for future designs.^{14,15,20} From the mechanistic point of view, the main feature of this chemical process is the presence of several successive steps, each one demanding acid-base catalysis. Thus, the catalyst must be structurally and electrostatically adapted to each transition state and intermediate in order to improve the kinetic of the overall reaction. The reaction mechanism we propose (Fig. 2) is based on our previous computational studies of the same reaction catalysed by two *de novo* protein designs, RA95.5-5 and RA95.5-8F.^{14,15} The resulting DFT/MM free energy profile is depicted in Fig. 3, where the one previously obtained for the same reaction catalyzed by RA95.5-8F *de novo* enzyme¹⁴ is also shown. As schematically shown in Fig. 2 (see Fig. S1 for atom labeling), the **first step** of the reaction involves the attack to the carbonyl carbon of the substrate by nitrogen of the Lys93:H (N_{K93}) leading to the formation of a metastable zwitterion I_1 . This intermediate is mainly stabilized by the interaction of the partial charge formed on the N_{K93} with the aromatic ring of the Phe98:L, with the Trp103:H, and through an hydrogen bond interaction between O_1 and the hydroxyl group of Tyr36:L. In the **second step**, the hydrogenation of the carbonyl oxygen of the substrate to form the alcohol takes place. The formation of the intermediate I_2 is characterized by the presence of an alkoxide in Tyr36:L which interacts with one of the hydrogen atoms of the Lys93:H cation and with the Trp103:H amino group. The **third step** involves the deprotonation of the Lys93:H by the Tyr36:L leading to a carbinolamine intermediate (I_3). This step is strongly exergonic in both catalysts, being I_3 particularly stable in the CA (-18.1 kcal mol⁻¹), and with a relative activation free energy barrier of just 5.5 kcal mol⁻¹. This hydrogen transfer is associated with the formation of the N_{K93} -C₂ bond, along with the recovery of the neutral charge which appears to play a key role in the stability of the carbinolamide. The **fourth step** corresponds to the formation of the first product of the reaction, the 6-methoxy-2-naphthylacetic acid (6-MNA), and an enamine intermediate (I_4).

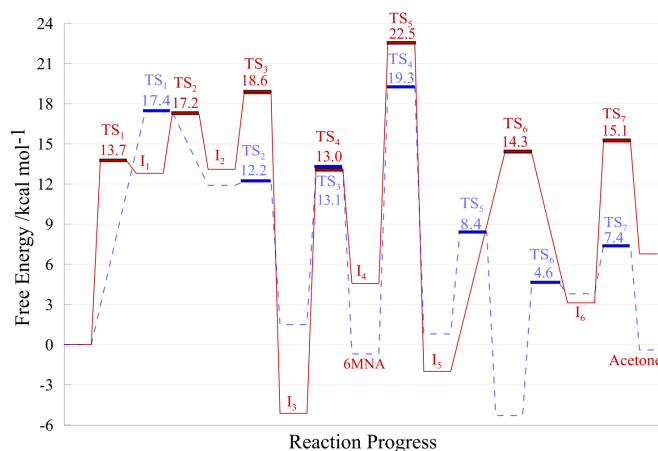


Fig. 3 M062X/6-31+G(d,p):AM1/MM free energy profile for the reaction mechanism of the Retro-Aldol reaction of the (R)-methodol enantiomer catalyzed by catalytic antibody 33F12 (red line) and by RA95.5-8F (blue line). RA95.5-8F profile is derived from ref. 14.

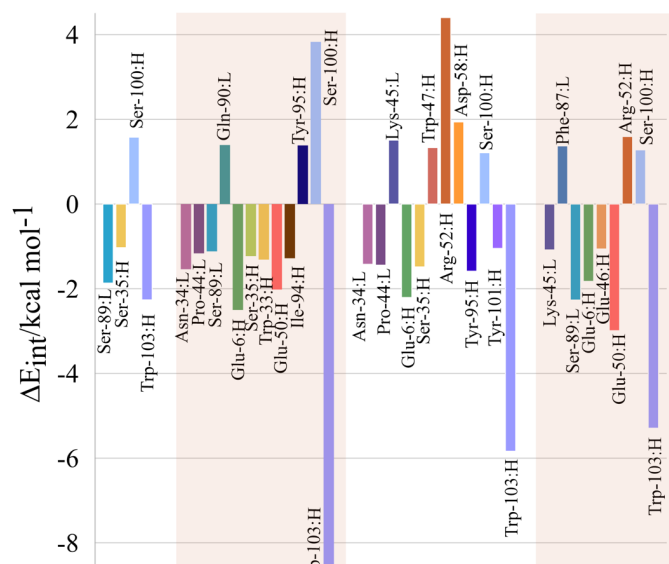


Fig. 4 Averaged substrate-protein activation interaction energies for the key steps of retro-aldol reaction catalyzed by 33F12. Only are displayed those residues showing interaction energies higher than 1 kcal mol⁻¹ (in absolute value).

The reaction involves the deprotonation of the hydroxyl group of the substrate (O₂) by Tyr36:L, the breaking of the C₃-C₄ bond of the substrate and the formation of a water molecule. The analysis of the geometries suggests that the transition state is assisted by the residues Trp103:H, Phe98:L and Asn34:L. This step has been traditionally the rate-determining step of the *de novo* enzymes^{15,21} with the exception of the RA95.5-8F^{14,19,22}, whose catalytic tetrad creates a hydrogen-bonding network able to enhance the kinetic of this step. The high value observed in the energy barrier for the CA (18.6 kcal mol⁻¹) could be ascribed to the absence of this organized network of polar residues in the active site. In the **fifth step**, the proton on Tyr36:L is transferred to the nucleophilic sp² carbon atom (C₃) of the enamine compound, to form the intermediate iminium Schiff base (I₅). The stability of the intermediate I₅ is due to the electrostatic interaction of the Schiff base with Asn34:L and Ser100:H, and the interaction of the negative charged Tyr36:L with Trp103:H. This is the rate-determining step of the overall reaction in both the CA and the RA95.5-8F *de novo* enzyme, although the catalytic efficiency of the latter is clearly higher than that found for the CA. The free energy value obtained for this step in the CA is 27.6 kcal mol⁻¹, which appear to be slightly overestimated by comparison with the experimental value (21.5 kcal mol⁻¹).²³ Nevertheless, our value is computed with respect intermediate I₃, which can be overstabilized. The energy difference between the rate limiting TS₅ and reactants is 22.5 kcal mol⁻¹, which would be in very good agreement with the experimental data. The nucleophilic attack of the water molecule and the reestablishment of the electroneutrality takes place in the **sixth step**, leading to the formation of the carbinolamine species (intermediate I₆). In the **seventh step**, the Tyr36:L behaves, like in TS₃, as both acid and basic species: it transfers a proton to the Lys93:H, promoting the breaking of the N_{K93}-C₂ bond; and withdraws the hydrogen from the hydroxyl group (O₁), yielding the acetone molecule, and

restoring the CA. The last transition state is stabilized by interactions with Trp103:H.

The comparative analysis between the *de novo* RA95.5-8F and the CA 33F12 shows how the former requires less chemical steps than the later. Thus, the CA needs two steps to reach I₂ intermediate while the same transformation is concerted in the RA95.5-8F enzyme. Nevertheless, the activation free energy of both systems only differs 0.2 kcal mol⁻¹, and I₁ in the CA is not kinetically relevant (according to the low energy barrier for the reverse decomposition into reactants). The opposite behaviour is observed in the transformation from the Schiff base, I₅, to the carbinolamine I₆ intermediate; the CA catalyzes this transformation in one single step while in the *de novo* enzyme it requires two step. It is remarkable nevertheless, how the former is clearly more efficient than the CA, according to the lower energy barriers. Overall, the reaction is an endergonic process in the CA (the product energy value is 6.7 kcal mol⁻¹ higher than the reactants), while it is almost thermoneutral in the *de novo* RA95.5-8F^{14,15} enzyme. The activation interaction energy decomposed by the contribution of each residue for the key chemical steps are shown in Figure 4 (see Supporting Information for details). These contributions are expressed as the difference of the electrostatic and Lennard-Jones interaction energies between the substrate and the protein, per residue, between each TS and its preceding intermediate. Analysis of these plots allows identifying and confirming the role that each amino acid plays in the catalytic process. As can be seen, the first step of the reaction in the CA (TS₁) is stabilized by the interaction with the residues Ser89:L, Ser35:H and Trp103:H, in opposition to the Ser100:H residue which disfavors/destabilizes the conversion. The second step (TS₂) is mainly favoured by the interaction with the residue Trp103:H, which plays a predominant role in this step and, to a lesser extent, by the residues Glu6:H and Glu50:H. On the other hand, a quite strong unfavourable interaction with Ser100:H is still present. The favourable interaction with the Trp103:H remains in the fourth step (TS₄), accompanied by a strong effect of the Glu6:H. Arg52:H together with Asp58:H and Ser100:H, clearly disfavour the reaction by stabilizing the intermediate I₃ with respect to TS₄. Finally, several unfavourable interactions can be identified for the fifth step (TS₅), such as Phe87:L, Arg52:H and Ser100:H, responsible for stabilizing the intermediate I₄ with respect to TS₅. Trp103:H, which assisted the reaction in the first

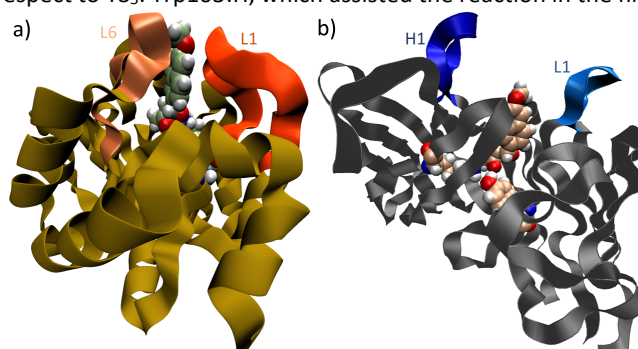


Fig. 5. a) *de novo* RA95.5-8F, in dark orange loop L1 and in sandy brown loop L6. b) Catalytic antibody 33F12, loop H1 is in dark blue, and loop L1 in light blue. The size of the ribbons is proportional to their mobility.

steps, together with Glu6:H, Glu50:H and Ser89:L stabilize the transition state. The evaluation of these results with the previously reported for the *de novo* RA95.5-8F enzyme¹⁴, allows to identify the role played by key residue in both systems. Interestingly, the comparison of the RMSD for the loops located in the active site (see Fig. 5) shows less movement in the CA than in the *de novo* enzyme throughout the key steps of the reaction (Fig. S10). This factor, together with the greater accessibility of the enzyme active site, entails a more favorable formation and removal of the product 6-MNA in the later. In addition, the deeply buried hydrophobic pocket of the 33F12, provides more stability to the neutrally charged intermediate I₃, the carbinolamine, that in turn provokes higher energy barrier for its decomposition.

In conclusion, the exploration of the retro-aldolase reaction catalysed by the CA 33F12 Fab' by means of QM/MM MD simulations allowed obtaining the complete free energy landscape and determining the molecular mechanism of the complex multi-step reaction. Analysis of electrostatic and Lennard-Jones interactions between the protein and the substrate in the key states of the process sheds light on the role played by the different residues of the active site. The obtained reaction mechanism is comparable to the one previously computed for the same reaction catalysed by the *de novo* design RA95.5-8F, but with noticeable differences. Thus, the first step in the *de novo* enzyme takes place in two steps in the CA catalysed reaction, while the opposite scenario is observed in the formation of the Schiff base from the protonation of the enamine, which takes place in two steps in the *de novo* enzyme and in a concerted manner in the CA. In fact, this is the rate limiting step of the full retro-aldol reaction catalysed by the CA. The activation free energy of this step appears to be slightly overestimated, by comparison with the available experimental data. Nevertheless, the lower catalytic efficiency of this biocatalyst, by comparison with the *de novo* design RA95.5-8F, reproduces the experimental kinetic trend. Our results show how Trp103:H, Glu50:H and Glu6:H are crucial to stabilize the TSs of the three key steps in the reaction. The hydrophobic microenvironment accounts for the pK_A of the ε-amino group of the active site Lys93:H,¹⁰ as in the *de novo* enzyme.²⁴ This confirms similar strategies adopted by both kind of catalysts. This work is a dowel in the study of the retro-aldol catalysts, and thereof it is possible to deduce information to propose new mutations able to improve *de novo* designed enzymes.

Acknowledgements

We would like to thank the Spanish Ministerio de Ciencia e Innovación (PGC2018-094852-B-C21), Generalitat Valenciana (AICO/2019/195) and Universitat Jaume I (UJI-B2020-03). DDR thanks the Generalitat Valenciana for a Grisolia PhD grant (GRISOLIAP/2016052). Authors acknowledge computational resources from the Servei d'Informàtica of Universitat Jaume I.

References

1 R. A. Lerner and A. Tramontano, *Sci. Am.*, 1988, **258**, 58–

70.

2 R. A. Lerner, S. J. Benkovic and P. G. Schultz, *Science*, 1991, **252**, 659–667.

3 P. G. Schultz, *Angew. Chem. Int. Ed. Engl.*, 1989, **18**, 1283–1295.

4 C. Gustafsson, S. Govindarajan and R. Emig, *J. Mol. Recognit.*, 2001, **14**, 308–314.

5 Z. Shao and F. H. Arnold, *Curr. Opin. Struct. Biol.*, 1996, **6**, 513–518.

6 S. Martí, J. Andrés, E. Silla, V. Moliner, Tuñón and J. Bertran, *Angew. Chemie - Int. Ed.*, 2006, **46**, 286–290.

7 S. Raza, L. Fransson and K. Hult, *Protein Sci.*, 2001, **10**, 329–38.

8 R. C. Nowiruki, M. R. Tam, L. C. Golrhtein, L. Stong, C. C. Kuo, L. Corey, W. E. Stamm and H. H. Hanrhfield, *Biotechnol. Biol. Front.*, 2019, 57–72.

9 F. R. Seiler, P. Gronski, R. Kurrle, G. Lüben, H. -P Harthus, W. Ax, K. Bosslet and H. -G Schwick, *Angew. Chemie Int. Ed. English*, 1985, **24**, 139–160.

10 J. Wagner, R. a Lerner and C. F. Barbas, *Science*, 1995, **270**, 1797–1800.

11 C. F. Barbas III, A. Heine, G. Zhong, T. Hoffmann, S. Gramatikova, R. Björnstedt, B. List, J. Anderson, E. A. Stura, I. A. Wilson and R. A. Lerner, *Science*, 1997, **278**, 2085–2092.

12 D. Hilvert, *Annu. Rev. Biochem.*, 2000, **69**, 751–93.

13 X. Zhu, F. Tanaka, R. A. Lerner, C. F. Barbas and I. A. Wilson, *J. Am. Chem. Soc.*, 2009, **131**, 18206–18207.

14 D. De Raffeale, S. Martí and V. Moliner, *ACS Catal.*, 2019, **9**, 2482–2492.

15 D. De Raffeale, S. Martí and V. Moliner, *ACS Catal.*, 2020, **10**, 7871–7883.

16 L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Röthlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas, D. Hilvert, K. N. Houk, B. L. Stoddard and D. Baker, *Science*, 2008, **319**, 1387–1391.

17 R. Obexer, A. Godina, X. Garrabou, P. R. E. Mittl, D. Baker, A. D. Griffiths and D. Hilvert, *Nat. Chem.*, 2016, **9**, 50–56.

18 A. Romero-rivera, *ACS Catal.*, 2017, **7**, 8524–8532.

19 J. W. Schafer, I. Zoi, D. Antoniou and S. D. Schwartz, *jacs*, 2019, **141**, 10431–10439.

20 P. G. Schultz and R. A. Lerner, *Science*, 1995, **269**, 1835–1842.

21 S. Bjelic, Y. Kipnis, L. Wang, Z. Pianowski, L. Wang, Z. Pianowski, G. Kornhaber, M. Su, J. Seetharaman, R. Xiao, G. Kornhaber, J. F. Hunt, L. Tong, D. Hilvert and D. Baker, *J. Mol. Biol.*, 2014, **426**, 256–271.

22 R. Obexer, A. Godina, X. Garrabou, P. R. E. Mittl, D. Baker, A. D. Griffiths and D. Hilvert, *Nat. Chem.*, 2017, **9**, 50–56.

23 F. Tanaka, R. A. Lerner, C. F. Barbas, N. Torrey, P. Road and L. Jolla, *J. Am. Chem. SOC.*, 2000, **122**, 4835–4836.

24 J. K. Lassila, D. Baker and D. Herschlag, *Proc. Natl. Acad. Sci.*, 2010, **107**, 4937–4942.

Conflicts of interest

There are no conflicts to declare.