

Títol del treball:

**EFFECTE DE LES MUTACIONS EN LA VARIANT RA95.5-8F
SOBRE L'ACTIVITAT RETRO-ALDÒLICA**

Estudiant: Cristina Duran i Rebenaque

Grau en Química

Correu electrònic: cristina9d@gmail.com

Tutor: Sílvia Osuna Oliveras

Cotutor*:

Empresa / institució: Universitat de Girona

Vistiplau tutor (i cotutor*):

Nom del tutor: Sílvia Osuna Oliveras

Nom del cotutor*:

Empresa / institució: Universitat de Girona

Correu(s) electrònic(s): silvia.osuna@udg.edu

*si hi ha un cotutor assignat

Data de dipòsit de la memòria a secretaria de coordinació:

Resum

Els enzims, que augmenten la velocitat de les reaccions químiques molts ordres de magnitud, són els catalitzadors més eficients de la Terra. Per aquest motiu, i per totes les seves característiques avantatjoses, estan esdevenint sistemes molt atractius per a les indústries. A hores d'ara, però, la biocatàlisi aplicada a aquest camp encara és bastant escassa degut a algunes limitacions que presenten els enzims. Per conèixer, millorar i superar aquestes limitacions, s'utilitza el disseny d'enzims. Aquesta tècnica, experimental o computacional, sempre segueix el mateix procés. L'evolució dirigida és l'estratègia existent més potent pel disseny d'enzims. Aquesta tècnica ha estat molt utilitzada, per exemple pel disseny de la retro-aldolasa 95 (RA95), però presenta les principals limitacions de no ser racional i d'alt cost econòmic.

La RA95 promou el trencament d'enllaços carboni-carboni, partint d'un aldol, obtenint el corresponent aldehid i una cetona. Aquesta reacció s'anomena retro-aldòlica i concretament promou el trencament del metodol a 6-metoxi-2-naftaldehid i acetona. Segueix una ruta de reacció de diversos passos, un dels quals és la formació d'un adducte covalent entre el substrat i la lisina catalítica (base de Schiff). La RA95 ha anat patint una sèrie de mutacions, donant lloc a diferents variants de l'enzim. L'última, i la més eficient, és la RA95.5-8F, que presenta una tètrada catalítica. Aquesta prové d'aplicar varies mutacions a la RA95.5-8 tant al centre catalític com a la perifèria de l'enzim.

L'objectiu d'aquest estudi és analitzar i conèixer la contribució sobre l'activitat catalítica dels residus mutats que no es troben al centre catalític. Mitjançant el "Shortest Path Map" (SPM) aplicat sobre la RA95.5-8, es poden predir els residus importants a la dinàmica conformacional de l'enzim. Així doncs, partint de l'última variant, s'hi reverteixen els residus adients (pels de la RA95.5-8). Es segueix un protocol de simulació de dinàmica molecular (MD) sobre el pas de reacció posterior a la formació de la base de Schiff. Això permet l'anàlisi de la distància catalítica i la "*Root Mean Square Deviation*" (RMSD). A més a més, s'utilitza el programa PyMol per poder visualitzar millor el comportament de l'enzim al llarg de la simulació. Els resultats de les simulacions realitzades per cadascuna de les variants indiquen que l'eina computacional SPM és un bon mètode per determinar quines posicions de l'enzim són importants per l'activitat retro-aldòlica.

Resumen

Las enzimas, que aumentan la velocidad de las reacciones químicas muchas órdenes de magnitud, son los catalizadores más eficientes de la Tierra. Por eso, y por todas sus características ventajosas, están siendo muy atractivos para las industrias. Aún así, hoy en día, la biocatálisis aplicada a este campo es todavía bastante escasa debido a algunas limitaciones que presentan las enzimas. Para conocer, mejorar y superar estas limitaciones, se utiliza el diseño de enzimas. Esta técnica, experimental o computacional, siempre sigue el mismo proceso. La evolución dirigida es la estrategia existente más potente para el diseño de enzimas. Esta técnica está siendo muy utilizada, por ejemplo para el diseño de la retroaldolasa 95 (RA95) pero presenta las principales limitaciones de no ser racional i de alto valor econòmico.

La RA95 promueve la rotura de enlaces carbono-carbono, partiendo de un aldol, obteniendo el correspondiente aldehído y una cetona. Esta reacción se denomina retro-aldólica y concretamente promueve la rotura del metodol a 6-metoxi-2-naftaldehido y acetona. Sigue una ruta de reacción de varios pasos, uno de los cuales es la formación de un aducto covalente entre el sustrato y la lisina catalítica (base de Schiff). La RA95 ha ido sufriendo una serie de mutaciones, dando lugar a diferentes variantes de la enzima. La última, y la más eficiente, es la RA95.5-8F, que presenta una tétrada catalítica. Esta proviene de aplicar varias mutaciones en la RA95.5-8 tanto al centro catalítico como en la periferia de la enzima.

El objetivo de este estudio es analizar y conocer la contribución sobre la actividad catalítica de los residuos mutados que no se encuentran en el centro catalítico. Mediante el "Shortest Path Map" (SPM) aplicado sobre la RA95.5-8, se pueden predecir los residuos más relevantes a la dinámica conformacional de la enzima. Así pues, se parte de la última variante y se revierten los residuos adecuados (por los de la RA95.5-8). Se sigue un protocolo de simulación de dinámica molecular (MD) sobre el paso de reacción posterior a la formación de la base de Schiff. Esto permite el análisis de la distancia catalítica y la "*Root Mean Square Deviation*" (RMSD). Además, se utiliza el programa PyMol para visualizar mejor el comportamiento de la enzima durante la simulación. Los resultados de las simulaciones realizadas por cada una de las variantes indican que la técnica computacional SPM es un buen método para determinar qué posiciones de la enzima son importantes para la actividad retroaldólica.

Abstract

Enzymes, which increase the chemical reactions rate many orders of magnitude, are the most efficient catalysts on Earth. For this reason, and for all their advantageous features, they are becoming very attractive for the industries. However, biocatalyst applied to this field, at this time, is still quite scarce due to some enzymes limitations. These are some barriers that must be overcome. To understand, improve and overcome these barriers, enzyme design is used. This technique can be both experimental and computational, but it always follows the same process. The directed evolution is the most powerful strategy, of the three existing ones, for the enzymes design. This technique has been extensively used, for example for the design of the retro-aldolase 95 (RA95).

RA95 promotes the breakdown of carbon-carbon bonds, starting from an aldol and obtaining the corresponding aldehyde and ketone. This reaction is called retro-aldol reaction and specifically promotes the breakdown of the methodol to 6-methoxy-2-naphthaldehyde and acetone. It follows a reaction path of several steps, one of which is the formation of a covalent adduct between the substrate and the catalytic lysine (Schiff base). Several mutations have been introduced into RA95, leading to different variants of the enzyme. The last one, and the most efficient, is RA95.5-8F, which has a catalytic tetrad. This new variant is originated applying some mutations to RA95.5-8 both in the catalytic centre and in the periphery of the enzyme.

The aim of this study is to analyse and understand the contribution on catalytic activity of mutated residues that are not in the catalytic centre. Using the "Shortest Path Map" (SPM) tool applied to RA95.5-8, the most relevant residues can be predicted on the basis of the conformational dynamics of the enzyme. Thus, several mutations in RA95.5-8F are reverted back to the original residue found for the RA95.5-8 variant. A molecular dynamics (MD) simulation protocol is followed based on the reaction intermediate formed after the formation of the Schiff base. This allows the rationalization of how the catalytic distance and the "Root Mean Square Deviation" (RMSD) are affected after reverting the selected mutations. In addition, the PyMol program is used for a better visualization of the behaviour of the enzyme during the simulation. The results of the simulations for each of the variants show that the SPM computational tool is a great method to determine which positions of the enzyme are relevant for the retro-aldol activity.

Índex

1. Introducció	2
1.1. Reacció retro-aldòlica.....	2
1.2. Retro-aldolasa 95.....	3
1.2.1. RA95.5-8F.....	5
1.2.2. RA95.5-8.....	6
2. Objectives	7
3. Metodologia	8
3.1. SPM (“Shortest Path Map”).....	8
3.2. Preparació de la proteïna. Reversió de mutacions.....	8
3.3. Preparació del sistema.....	9
3.4. Dinàmica molecular (MD).....	10
3.5. Anàlisi dels resultats.....	10
3.5.1. Distància catalítica.....	10
3.5.2. RMSD de la base de Schiff.....	11
3.5.3. Visualització de la trajectòria.....	12
3.6. Part ètica i sostenible de l’estudi.....	12
4. Resultats i discussió	13
4.1. “Variant 75”: RA95.5-8F amb la posició 75 revertida.....	14
4.2. “Variant 90”: RA95.5-8F amb la posició 90 revertida.....	16
4.3. “Variant 135”: RA95.5-8F amb la posició 135 revertida.....	18
4.4. “Variant 151”: RA95.5-8F amb la posició 151 revertida.....	20
4.5. “Variant 75_135”: RA95.5-8F amb les posicions 75 i 135 revertides.....	22
4.6. “Variant 90_151”: RA95.5-8F amb les posicions 90 i 151 revertides.....	24
4.7. Comportament general.....	26
5. Conclusions	27
6. Bibliografia	28

1. Introducció

Els enzims són biomolècules, que acceleren reaccions químiques en molts ordres de magnitud. Aquest increment de la velocitat de reacció és degut a la disminució de l'energia d'activació. Això fa que els enzims siguin els catalitzadors més eficients que es coneixen, a més a més són molt importants per la vida a la Terra. La gran majoria de processos metabòlics que tenen lloc dins les cèl·lules involucren enzims, assolint les velocitats de reacció necessàries per mantenir l'organisme viu. Generalment, són molt específics i selectius pel que fa al substrat, sobretot degut a la càrrega, la forma i les característiques hidrofíliques o hidrofòbiques del centre catalític. Això proporciona que el rendiment del producte desitjat sigui elevat. A més a més, no són consumits durant la reacció, no alteren l'equilibri químic, són biodegradables i no són tòxics. Totes aquestes característiques avantatjoses fan que els enzims siguin catalitzadors molt atractius per les indústries. Es podrien utilitzar per convertir els processos industrials en beneficiosos des del punt de vista mediambiental i també reduir costos de fabricació.

Tot i aquests avantatges, la biocatàlisi a nivell industrial encara és bastant escassa. Això es deu a algunes limitacions com són l'estabilitat dels enzims naturals sota les condicions requerides (temperatura, pH o solvent), la manca d'enzims naturals que catalitzin reaccions concretes o, si aquests existeixen, que no estiguin optimitzats pel substrat d'interès industrial.

Pel disseny d'enzims hi ha tècniques tant experimentals com computacionals, però totes estan basades en el mateix procés: 1) selecció dels punts de mutació, 2) aplicació de les mutacions, i 3) avaluació de les noves variants de l'enzim obtingudes. Hi ha tres estratègies principals pel disseny d'enzims. Es presenten a continuació.

- *No racional*. Generació d'un elevat nombre de variants al laboratori amb mutacions aleatòries. Les noves variants són actives però no es sap perquè ho són.
- *Semi-racional*. Combinació de la generació d'una sèrie de variants predites mitjançant càlculs, on posteriorment s'aplicarà la tècnica d'evolució dirigida (DE) al laboratori per tal de millorar-ne l'activitat.
- *Racional*. Limitació del procés a un nombre reduït de variants, normalment predites utilitzant eines computacionals. S'obté un gran coneixement sobre l'efecte de les mutacions introduïdes.

L'evolució dirigida ha esdevingut la tècnica no racional més potent per a la producció d'enzims per a qualsevol procés. Mitjançant aquesta tècnica s'han dissenyat gran quantitat d'enzim^[8]. Un exemple n'és la retro-aldolasa 95 (RA95), tipus d'aldolasa classe I que, com a tal, promou el trencament reversible d'enllaços carboni-carboni seguint una ruta de reacció de diversos passos^[5]. Aquesta ruta inclou, entre d'altres, la catàlisi mitjançant una amina (provinent d'una lisina) i la generació d'espècies intermèdies covalentment enllaçades anomenades bases de Schiff (formada per una lisina i el substrat corresponent). Concretament es tracta de reaccions retro-aldolòliques, catalitzades per les retro-aldolases.

1.1. Reacció retro-aldòlica

Una reacció aldòlica és una reacció de formació d'enllaços carboni-carboni. Aquesta es pot dividir en dues parts. Primerament té lloc l'addició aldòlica, que involucra l'addició nucleòfila de l'enolat d'una cetona a un aldehid, formant una β -hidroxicetona o aldol. Sovint l'aldol pateix una condensació aldòlica, en la qual el producte perd una aigua i es forma una cetona α,β -insaturada.

Una reacció retro-aldòlica, pot fer també la ruta de trencament dels enllaços carboni-carboni (reacció inversa a l'addició aldòlica). Es parteix d'un aldol o β -hidroxicetona i, degut a l'actuació d'un enzim que conté una amina, té lloc un trencament de l'enllaç carboni-carboni, obtenint una cetona i un aldehyd. En aquest cas, concretament, es promou el trencament del 4-hidroxi-4-(6-metoxi-2-naftil)-2-butanona (comunament anomenat metodol) a 6-metoxi-2-naftaldehid i acetona, utilitzant la RA95 com a enzim (veure *Figura 1*).

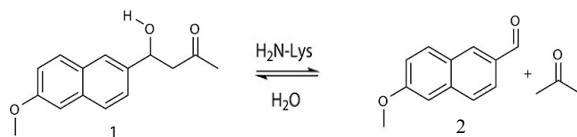


Figura 1. Reacció global del metodol (1) donant el corresponent naftaldehid (2) i acetona.

Com s'ha esmentat aquesta reacció inclou diferents passos de reacció. Un d'aquests és la catalisi mitjançant una amina (lisina, Lys en codi de tres lletres), que genera diferents espècies intermèdies de l'adducte covalent que formen una lisina catalítica i el substrat, el que s'anomena base de Schiff. Una base de Schiff es defineix com a grup funcional que conté un doble enllaç carboni-nitrogen, veure *Figura 2*.

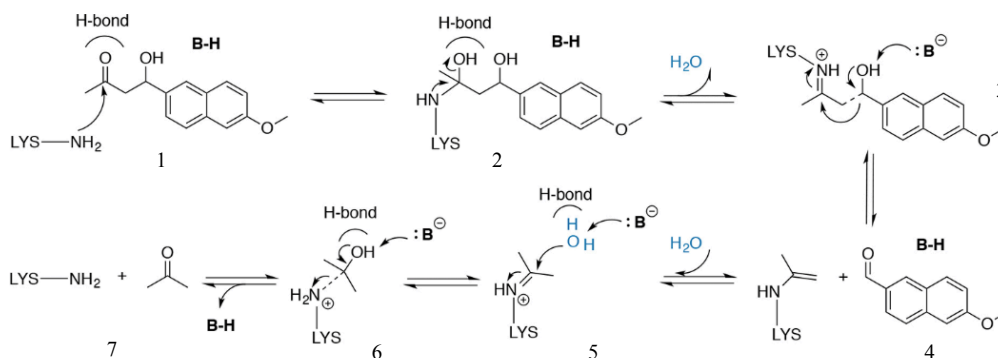


Figura 2. Mecanisme de reacció de la catalisi mitjançant una amina (Lys) en una reacció retro-aldòlica. Es parteix del substrat (1) per formar la base de Schiff (3) i finalment, els productes (4 i 7).

1.2. Retro-aldolasa 95

L'enzim artificial generat mitjançant tècniques racionals de disseny d'enzims és la RA95^[2]. Aquest és l'encarregat d'accelerar la reacció retro-aldòlica. Com s'ha esmentat abans, és una aldolasa tipus I, que promou el trencament reversible de enllaços carboni-carboni. Aquest enzim, que presenta poca activitat retro-aldòlica ($k_{\text{cat}}/K_M=0.02-0.3 \text{ M}^{-1}\text{s}^{-1}$), es va evolucionar mitjançant l'evolució dirigida de manera que es van introduir una sèrie de mutacions no només al centre actiu, sinó també a les perifèries, per tal d'augmentar la seva activitat catalítica. Totes les variants van ser analitzades també computacionalment observant diferents posicions de l'intermedi corresponent en les simulacions de dinàmica molecular (MD) en un estudi realitzat al grup de recerca del CompBioLab^[7]. L'evolució que ha seguit es mostra a continuació:

- *RA95.0*: La primera variant de l'enzim, dissenyada amb el protocol computacional anomenat "inside-out"^[4], té 16 mutacions, 11 de les quals tenen lloc al centre actiu. Aquest conté una lisina a la posició 210 per facilitar el trencament de l'enllaç carboni-carboni mitjançant la

formació de la base de Schiff, un glutamat a la posició 53 per orientar la molècula d'aigua catalítica i una cadena hidrofòbica llarga per acomodar el substrat (anell naftalènic). Com es pot veure a la Taula 1, presenta un valor de $k_{cat}/K_M = 0.17 \text{ M}^{-1}\text{s}^{-1}$.

- *RA95.5*: És la primera variant evolucionada amb evolució dirigida^[5]. Conté cinc mutacions addicionals al centre actiu (V51Y, E53S, T83K, M180F i R182M) i una altra substitució a una distància lleugerament major d'aquest (D183N). L'aparició d'una segona lisina a la posició 83 és comuna a moltes aldolases classe I, la que promou un grup catiònic addicional al centre actiu per activar la Schiff base. A més a més, la nova tirosina a la posició 51 promou la desprotonació del β -alcohol. S'observen dues possibles orientacions per l'intermedi de la base de Schiff. Com es pot veure a la Taula 1, presenta un valor de $k_{cat}/K_M = 16 \text{ M}^{-1}\text{s}^{-1}$.
- *RA95.5-5*: Conté sis mutacions addicionals, tres de les quals estan a una distància petita del centre actiu (E53T, S110N i G178S) i tres altres que estan a la superfície de la proteïna (R23H, R43S i T95M). Les simulacions de dinàmica molecular realitzades a un estudi previ^[7] mostren com l'intermedi de la base de Schiff no està correctament situada per la catàlisi. Com es pot veure a la Taula 1, presenta un valor de $k_{cat}/K_M = 320 \text{ M}^{-1}\text{s}^{-1}$.
- *RA95.5-8*: Conté cinc mutacions addicionals, tres properes del centre actiu (K135N, S178V i G212D) i dues en una posició distal (S43R i F72Y). Presenta una tríada catalítica (Lys83, Tyr51 i Asn110). L'intermedi de la base de Schiff encara és lleugerament flexible. Com es pot veure a la Taula 1, presenta un valor de $k_{cat}/K_M = 1600 \text{ M}^{-1}\text{s}^{-1}$.
- *RA95.5-8F*: L'última variant de l'enzim, generada amb evolució dirigida^[5], és la que presenta una major activitat catalítica. Aquesta conté tretze noves modificacions. Inclou una nova tirosina al centre actiu (F180Y) i una leucina al lloc de la lisina catalítica original (K210L). Presenta una tètrada catalítica (Lys83, Tyr51, Asn110 i Tyr180), que està perfectament disposada per enllaçar el substrat i adaptar-se als canvis geomètrics i electrostàtics que tenen lloc durant els passos de reacció^[7]. Cal esmentar que aquesta variant presenta una activitat similar a les aldolases naturals de classe I. Com es pot veure a la Taula 1, presenta un valor de $k_{cat}/K_M = 34000 \text{ M}^{-1}\text{s}^{-1}$, 21 vegades més eficient que la variant anterior.

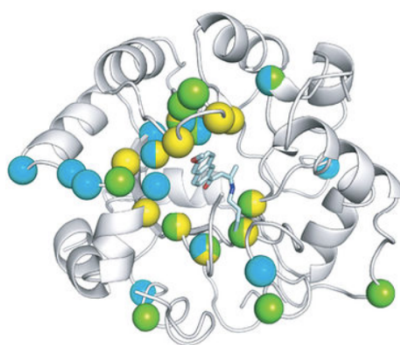


Figura 3. Mutacions de RA95 segons la variant. Mutacions de la RA95.0 en esferes grogues, de la RA95.5-8 en esferes verdes i de la RA95.5-8F en esferes de color blau.

A mesura que la RA95 ha anat evolucionant mitjançant la introducció de noves mutacions, ha anat millorant la seva activitat catalítica i, així mateix, la seva eficiència (k_{cat}/K_M). Per aquest motiu, l'última variant, la RA95.5-8F, serà la que mostrarà una eficiència òptima respecte totes les altres. A la Taula 1 es mostra l'eficiència i l'increment de l'activitat catalítica per a cada variant.

Taula 1. Eficiència i augment de l'eficiència (respecte la primera variant) per a cada variant de la RA95.

Variant	k_{cat}/K_M ($M^{-1}s^{-1}$)	Augment de l'eficiència
RA95.0	0.17	-
RA95.5	18	106
RA95.5-5	320	1880
RA95.5-8	1600	9400
RA95.5-8F	34000	200000

1.2.1. RA95.5-8F

L'última variant de la RA95 és la que presenta una major activitat catalítica i una major eficiència (veure Taula 1). La seva tètrada catalítica està perfectament situada per interaccionar amb el substrat i adaptar-se als canvis geomètrics i electrostàtics que tenen lloc durant la ruta de reacció. Aquesta està formada per la Lys83, Tyr51, Asn110 i Tyr180. Així doncs, la catàlisi mitjançant una amina serà provinent de la Lys83, ja que la lisina conté aquest grup amina per a dur a terme la reacció i la formació de la base de Schiff.

El mecanisme d'aquesta reacció retro-aldòlica utilitzant la tètrada catalítica de la RA95.5-8F com a centre catalític comença amb l'atac de la Lys83 al carbonil del metodol (substrat). Es forma un hemiaminal intermedi (un mateix carboni conté un grup hidroxil i una amina). Aquest allibera l'hidroxil en forma d'aigua i s'obté la corresponent base de Schiff protonada. L'ió imini actua com a donador d'electrons per afavorir el trencament de l'enllaç carboni-carboni i formar els corresponents aldehid i enamina. La Tyr180 desprotonada, està correctament posicionada per actuar com a base catalítica. Seguidament es produeix una protonació i una hidròlisi de l'enamina per acabar obtenint l'acetona final i l'enzim regenerat. (Veure *Figura 1* i *Figura 4*)

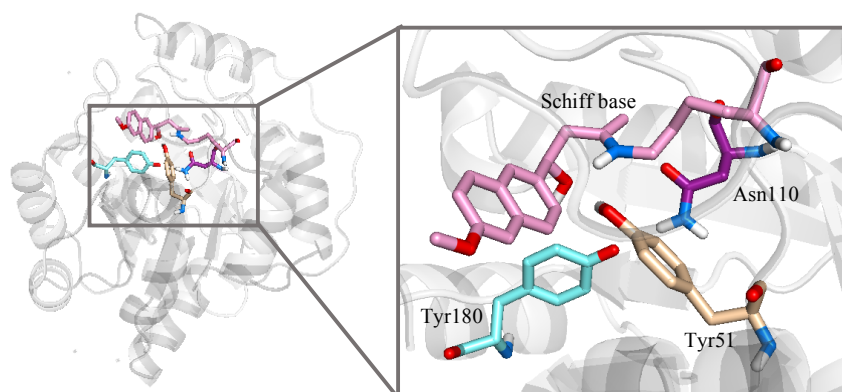


Figura 4. Tètrada catalítica de la RA95.5-8F. En rosa l'intermedi de la base de Schiff (adducte entre la Lys83 i el substrat), en cian la Tyr180, en beix la Tyr51 i en lila la Asn110.

1.2.2. RA95.5-8

La variant que presenta la segona millor eficiència és la RA95.5-8 ($k_{cat}/K_M = 1600 \text{ M}^{-1}\text{s}^{-1}$, veure Taula 1). Aquesta pateix 13 modificacions per a convertir-se en la última variant, la RA95.5-8F. En un estudi previ realitzat al grup de recerca CompBioLab^[7], es va demostrar com l'eina computacional anomenada "Shortest Path Map" (SPM) aplicat a la RA95.5-8 (veure metodologia), es podien predir aquestes mutacions. Aquesta eina, desenvolupada al CompBioLab, permet identificar quines posicions o residus tenen un paper més rellevant a la dinàmica conformacional de l'enzim. Cal esmentar que les 7 posicions predites amb el SPM coincideixen amb les posicions mutades anteriorment descrites. Dues de les mutacions predites es troben a una distància de 6 posicions respecte els SPM (Asn90 i Ser151), una a distància de 3 posicions (Asn135), una altra a una distància de 2 posicions (Arg75) i les dues restants a una distància de 1 posició (Ser214 i Arg216). Donat que experimentalment no es sap quina és la importància de les 13 mutacions introduïdes per a l'activitat retro-aldòlica de l'enzim, es va dissenyar aquest estudi per determinar com afecten les mutacions no incloses al SPM a l'activitat catalítica.

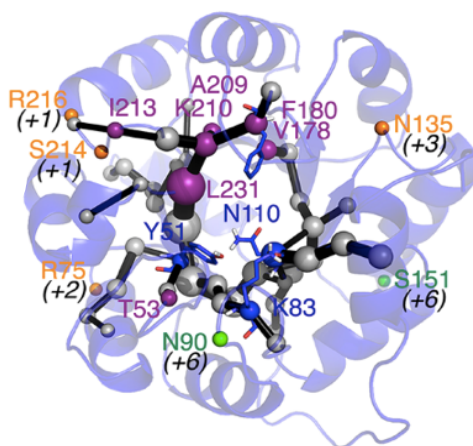


Figura 5. "Shortest Path Map" (SPM) de la variant RA95.5-8. En blau la tètrada catalítica, en lila les set mutacions predites pel SPM, en taronja les mutacions adjacents i en verd les mutacions que es troben desplaçades a +6 posicions en seqüència del conjunt de residus predits computacionalment.

2. Objectives

The man-made enzyme retro-aldolase (RA95), which catalyse the cleavage of the methodol substrate, has been evolved by means of Directed Evolution (DE) to present several computational mutations that improve its catalytic activity and so, its efficiency. Each variant contains at least 5 mutations located either at the active site or at distal positions. The last variant RA95.5-8F, the most efficient, contains a catalytic tetrad (Lys83, Tyr51, Asn110 and Tyr180), that is perfectly arranged to bind the substrate and promote the retro-aldol reaction. The previous less efficient variant RA95.5-8 contains a catalytic triad (Lys83, Tyr51 i Asn110). It is clear that the catalytic centre plays the most important role in the reaction mechanism, so the introduced mutations in the active site have a huge contribution on the catalytic improvement. However, each variant also contains mutations far away from the active site, which play a crucial role in improving the efficiency and stability of the enzyme.

The “Shortest Path Map” (SPM), developed at CompBioLab, is the perfect tool to predict or identify the most relevant residues to tune the conformational dynamics of the enzyme. The SPM of RA95.5-8 predicts mutations that have been introduced in this variant to become the RA95.5-8F variant. As seen in “Figura 5”, there are four residues located in adjacent positions of the SPM (Asn135, Arg75, Ser214 and Arg216) and two others at more than 5 positions far away in the sequence from the SPM (Asn90 and Ser151).

As mentioned before, retro-aldol reactions have a multistep reaction mechanism. It takes place *via* amine catalysis, where a covalent adduct between the substrate and the catalytic lysine is formed, known as Schiff base. In reference “Figura 2”, it is studied the third step, where the Schiff base has already been formed and is going to react with the base, in this case Tyr180.

The main goal of the current study is to analyse the effect for retro-aldol activity of those four mentioned mutations (D90N, G151S, P75R and E135N) that are not located in SPM, but in the proximities. The hypothesis is that those residues situated far away from the SPM predicted have a minor effect on the enzyme catalytic activity compared with those situated in adjacent positions. To accomplish this objective, a multistep computational procedure is proposed:

- The RA95.5-8F positions studied in each case will be reverted back to the RA95.5-8 residues. For each position or group of positions studied, new version of the enzyme is obtained.
- The structural features, configuration and dynamics of the different variants will be analysed by means of molecular dynamics (MD) simulations.
- The catalytic distance between the catalytic base (Tyr180) and the β -alcohol at the Schiff base intermediate will be analysed, as well as the flexibility of the latter intermediate with the “Root Mean Square Deviation” (RMSD). For a better visualization, PyMol program will be used.

3. Metodologia

3.1. SPM (“Shortest Path Map”)

Com s’ha esmentat, es parteix d’un estudi previ realitzat al grup de recerca CompBioLab^[7]. Per aquest, es va aplicar l’eina computacional anomenada “*Shortest Path Map*” (SPM) a la RA95.5-8. D’aquesta manera es podien predir quines mutacions d’aquest enzim revertir.

Aquesta eina permet identificar quines posicions o residus tenen un paper més rellevant a la dinàmica conformacional de l’enzim. Està basada en la utilització d’un algoritme anomenat Dijkstra. Aquest algoritme permet trobar quin és el camí més curt per anar des del primer residu fins l’últim de l’enzim. Per fer això, es mesura la distància entre parells de residus durant la simulació, amb els C α com a referència. Es considerarà que dos residus són propers en l’estructura tridimensional quan tinguin una distància menor a 5 Å durant tota la simulació. Una vegada seleccionats aquests residus, es calcula la seva correlació per poder establir si tenen un moviment comú o són independents entre ells. Per generar el SPM, només es té en compte aquells parells de residus que presentin una gran correlació, és a dir, que tinguin un moviment en comú i, a més a més, que siguin propers (distància menor a 5 Å). Tenint tots aquests factors en compte es descriu el camí espacial més curt des del primer fins l’últim residu, el SPM. S’utilitza PyMol per poder representar-lo en l’estructura proteica.

Encara que per aquest treball no s’utilitza aquesta eina, es parteix de l’estudi previ esmentat, en el qual aconseguia predir els residus importants de la RA95.5-8, descrits pel corresponent SPM (veure *Figura 5*). Així doncs, es fa una selecció dels residus que no es troben al SPM però que són propers a ell: Arg75, Asn135, Asn90 i Ser151.

3.2. Preparació de la proteïna. Reversió de mutacions

Es parteix de l’estructura RA95.5-8F cristal·litzada (codi PDB 4pa8) amb la Tyr180 desprotonada. Es volen revertir quatre mutacions que es troben a les proximitats del SPM, però no al camí de residus directament predits per l’eina computacional. Els residus més desplaçats respecte la seqüència del SPM (de més a menys) són: Arg75 (desplaçat 2 posicions), Asn135 (desplaçat 3 posicions), Asn90 (desplaçat 6 posicions) i Ser151 (desplaçat 6 posicions). Les reversions de cada posició es fan respecte els residus de la variant RA95.5-8, variant anterior a la RA95.5-8F. A part d’analitzar cada una de les posicions, també s’estudien combinacions entre elles.

Es va fer servir el programa PyMol^[3], per tal de poder superposar les dues variants de l’enzim que s’utilitzen. A més a més, PyMol proporciona la seqüència de cada una d’elles. Per a cada posició tractada, es canvia el residu de la RA95.5-8F pel residu de la variant RA95.5-8 que ocupa la mateixa posició. Així doncs, s’aconsegueix la última variant (amb la corresponent tètrada catalítica) però havent mutat alguns residus pels de la variant anterior (veure *Taula 2*).

D’aquesta manera s’obtenen tantes versions com posicions o combinacions entre elles es vulguin estudiar, totes en format PDB. Aquests fitxers proporcionen la seqüència de la proteïna i, així mateix, l’estructura en tres dimensions d’aquesta.

Taula 2. Mutacions a l'enzim RA95.5-8F.
Asp90 es muta a Asn90 (D90N)
Gly151 es muta a Ser151 (G151S)
Asp90 i Gly151 es muten a Asn90i Ser151 (D90N, G151S)
Pro75 es muta a Arg75 (P75R)
Glu135 es muta Asn135 (E135N)
Pro75 i Glu135 es muten a Arg75 i Asn135 (P75R, E135N)

3.3. Preparació del sistema

Una vegada obtingudes les variants resultants de les mutacions, es procedeix a la preparació del sistema. Aquest pas és essencial per a la posterior simulació de dinàmica molecular (MD). Els sistemes s'han simulat utilitzant el software AMBER16 amb el camp de forces ff99SB. Les simulacions s'han realitzat al clúster Galatea del CompBioLab, grup de recerca del IQCC, compost per 178 GTX1080 GPUs.

El terme *camp de forces* ("Force Fields") fa referència a la forma i paràmetres de les funcions matemàtiques que s'utilitzen per descriure l'energia potencial de l'enzim. Aquesta energia potencial té contribucions dels àtoms enllaçats ("bonded") i dels no-enllaçats ("nonbonded"). D'aquesta manera, l'energia potencial total es representa a l'equació [1]:

$$E_{potencial\ total} = E_{bonded} + E_{nonbonded} \quad [1]$$

Així mateix, les contribucions de les components d'energia del "bonded" [2] i del "nonbonded" [3] es representen en les següents addicions:

$$E_{bonded} = E_{enllaç} + E_{angle} + E_{diedre} \quad [2]$$

$$E_{nonbonded} = E_{electrostàtiques} + E_{Van\ der\ Waals} \quad [3]$$

La contribució dels àtoms enllaçats depèn de l'enllaç, l'angle entre els àtoms i l'angle diedre. Per altra banda, la component que prové dels àtoms no enllaçats depèn de les forces electrostàtiques i les forces de Van der Waals.

Per dur a terme la preparació del sistema s'utilitza TLEAP, un mòdul del programari AMBER. Es parteix dels fitxers PDB que s'obtenen a l'apartat anterior. Aquest mòdul addiciona els contraions necessaris per neutralitzar el sistema (Cl^- o Na^+) i una caixa solvatació cúbica d'aigua de 10 Å. Amb aquesta caixa de solvatació s'aconsegueix simular l'entorn de la proteïna amb aigües explícites.

La finalitat d'aquesta preparació del sistema és aconseguir els dos fitxers necessaris per la posterior dinàmica molecular. Un dels fitxers conté informació sobre el tipus d'àtoms i els paràmetres que s'utilitzen per calcular l'energia del sistema, el que s'anomena "tipologia" (fitxer PRMTOP). L'altre conté les coordenades del sistema (fitxer INPCRD). A més a més s'originen uns nous fitxers PDB de cada versió de l'enzim amb la caixa d'aigües per poder comprovar si el procediments anteriors s'han realitzat correctament.

3.4. Dinàmica molecular (MD)

Un cop obtinguts els fitxers amb la tipologia i les coordenades, s'inicia el protocol de simulació de dinàmica molecular.

- **Minimització:** es duu a terme en dos etapes. En la primera es minimitzen les posicions dels contraions i les molècules d'aigua de la caixa de solvatació, mentre es manté l'enzim fix. Cal apuntar que, per tal de mantenir fixos tots els àtoms dels enzims, s'utilitzen unes restriccions específiques. En la segona etapa, en canvi, es minimitza tot el sistema al complet per aconseguir una conformació inicial més estable.
- **Escalfament:** consisteix en escalfar progressivament el sistema de 0K a 300K. D'aquesta manera, es proporciona energia cinètica que dona peu al moviment dels àtoms del sistema primerament minimitzat. L'objectiu d'aquest procés és simular al màxim les condicions de laboratori durant la dinàmica. Com al cas de la minimització, també s'utilitzen restriccions momentàniament. Aquest procés es duu a terme a volum constant.
- **Equilibració:** aquest procés es realitza a pressió constant. El volum varia de manera que la densitat del solvent, aigua en aquest cas, es relaxa i s'adapta a les condicions del sistema (a una temperatura de 300K).

Després d'aquests tres passos previs, s'inicia la producció de la trajectòria de la dinàmica molecular. En aquest cas, el temps total de simulació és de 1000 ns. S'ha de tenir en compte que, per tal d'obtenir una major representativitat es fan 4 rèpliques de cada versió de l'enzim proposada. Això permet un major mostreig dels sistemes, ja que les velocitats inicials de la MD són aleatòries en cada rèplica, i per tant, permet aconseguir un major nombre de dades per poder analitzar estadísticament.

3.5. Anàlisi dels resultats

S'ha emprat el mòdul PTRAJ per analitzar els resultats de la dinàmica molecular. Aquest és un mòdul d'AMBER16 que permet obtenir informació important de la simulació. Té una gran varietat de funcions, per exemple, el simple anàlisi geomètric de coordenades, la conversió entre diferents formats de coordenades o l'anàlisi avançat de matrius i vectors. Tot i això, en aquest cas, s'estudien dos factors. Un d'ells és la distància catalítica i l'altre la RMSD ("*Root Mean Square deviation*") de la base de Schiff. A més a més s'utilitza el programa PyMol, una eina potent que permet visualitzar la trajectòria.

3.5.1. Distància catalítica

Com s'ha esmentat abans, la distància catalítica s'analitza amb el mòdul PTRAJ. Totes les simulacions de dinàmica molecular es fan sobre el pas 3 en referència a la *Figura 2*, per tant s'analitzen les conformacions que adopten els quatre residus del centre catalític mentre l'adducte covalent entre el substrat i la Lys83 està format. Cal recordar que el pas de reacció posterior a la formació de la base de Schiff és l'atac d'una base sobre el β -alcohol del substrat per a la seva desprotonació. La Tyr180, que forma part de la tètrada catalítica, té la funció de base que desprotona la base de Schiff. És per aquest motiu que la distància catalítica es descriu com la distància entre l'oxigen del β -alcohol de la base de Schiff i l'oxigen de la Tyr180 desprotonada, veure *Figura 6*.

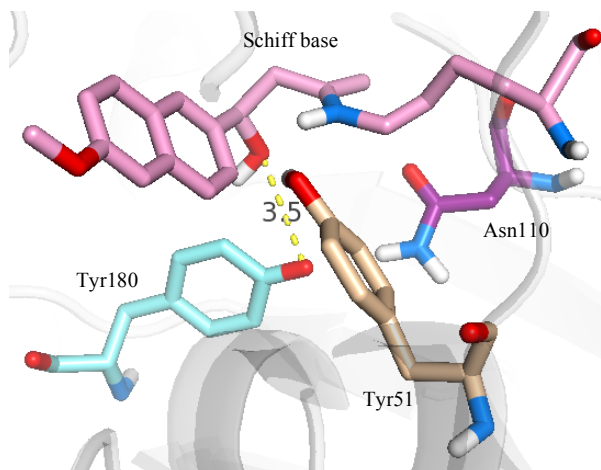


Figura 6. Distància catalítica (mesurada en Å). En rosa un intermedi de la base de Schiff (adducte entre la Lys83 i el substrat), en cian la Tyr180, en beix la Tyr51 i en lila la Asn110.

A la *Figura 6* s'hi exemplifica una distància catalítica, en aquest cas essent de 3.5 Å. Ara bé, el resultat de les simulacions de dinàmica molecular serà una trajectòria del sistema. Així doncs, no tindrem un sol valor de distància catalítica sinó una recopilació de la variació d'aquesta distància al llarg del temps de simulació, en concret, 1000 ns. Per poder analitzar la trajectòria s'utilitza el mòdul PTRAJ, que permet fer una gràfica d'aquesta distància catalítica en Àngstroms (eix y) enfront el temps en nanosegons (eix x).

Si el comportament d'aquesta gràfica és totalment constant, significa que la distància catalítica no ha variat en tota la simulació. Per contra, si té un comportament del tot irregular, equivaldria a moltes fluctuacions en el valor de la distància catalítica. A més a més es calcula la mitjana de les distàncies catalítiques de cada una de les rèpliques de totes les variants amb la seva corresponent desviació estàndard.

3.5.2. RMSD de la base de Schiff

El mòdul PTRAJ també proporciona el càlcul de la RMSD, el qual és una bona eina per estudiar la variació de la distància catalítica. Aquest càlcul mesura la desviació d'un conjunt de coordenades objecte d'estudi respecte un conjunt de coordenades de referència amb RMSD=0. Per aquest cas, s'estudia tan sols per la base de Schiff, no per la proteïna total. Aquest càlcul ve descrit per l'equació [4].

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=0}^N (X_i - Y_i)^2} \quad [4]$$

on N és el nombre d'àtoms; X_i és el vector de les coordenades de l'àtom i que es vol estudiar; i Y_i és el vector de les coordenades de l'àtom i de referència.

Així doncs, un valor de RMSD igual a 0 indicarà que no hi ha desviació del conjunt de coordenades estudiat respecte al de referència i que, per tant, hi haurà una superposició perfecte. Per contra, si el valor de RMSD és gran, indicarà que hi ha desviació i que, conseqüentment, no hi ha superposició.

Aplicat al cas de l'estudi en qüestió, s'estudia la desviació del coordenades de la base de Schiff. El mòdul PTRAJ permet fer una gràfica d'aquesta variació, és a dir la RMSD (eix y) enfront el temps en nanosegons (eix x). Si la RMSD té un valor relativament petit al llarg del temps, indica que la base de Schiff no presenta una gran variació de la posició dels seus àtoms respecte els de referència. Ara bé, si el valor de RMSD és major, significa que la base de Schiff té més moviment respecte els àtoms de referència.

Ara bé també s'ha de tenir en compte si el valor de RMSD es manté constant o no al llarg del temps. Si és així, indicarà que la variació de la distància catalítica durant el temps de simulació no és massa gran, independentment de la seva posició respecte els àtoms de referència. D'aquesta manera es podrà intuir una certa superposició dels residus. Si, per contra, no es manté constant, no hi haurà tanta superposició entre ells. Aquest valor estarà estretament relacionat amb la desviació estàndard. Un valor constant de RMSD indicarà, un valor petit de desviació estàndard.

3.5.3. Visualització de la trajectòria

Per poder entendre d'una manera més clara i visual el comportament de les variants estudiades, s'utilitza el programa PyMol. Aquest permet visualitzar la trajectòria obtinguda després de realitzar la simulació de dinàmica molecular.

S'obtenen una sèrie de fitxers en format PDB que contenen la informació dels moviments i conformacions que adopta la RA durant el temps de simulació. El punt d'interès d'aquest estudi és el centre catalític, així que s'enfoca la visualització en aquest. El que s'aconsegueix veure mitjançant aquest programa és la superposició de les diferents conformacions i posicions que va adoptar cada residu de la tètrada catalítica durant els 1000 ns que dura la simulació. Així doncs, facilita la mateixa informació que proporcionaria un vídeo, però sense moviment.

La imatge resultant mitjançant PyMol va estretament relacionada amb les gràfiques que s'obtenen en els dos apartats anteriors. Així doncs, si una rèplica té una RMSD constant al llarg del temps, es veurà una gran superposició de la base de Schiff. Per contra, si té un valor variant, els residus adoptaran diferents conformacions durant la simulació. Pel que fa a la distància catalítica, es pot analitzar d'una manera més visual.

3.6. Part ètica i sostenible de l'estudi

En l'estudi de la RA95.5-8F s'ha utilitzat tant sols eines computacionals, les quals permeten fer un estudi acurat de l'enzim en qüestió. Així doncs no requereix cap part experimental, fet que implica no utilitzar reactius o solvents, que podrien ser perjudicials pel medi ambient. Per aquest motiu, es considera que l'estudi és respectable amb el medi ambient.

A més a més, les dades presentades al treball no han sigut copiades ni canviades per tal d'aconseguir millors resultats. Tota la bibliografia utilitzada s'ha enumerat i detallat a l'apartat corresponent. S'ha evitat plagi.

4. Resultats i discussió

Després d'haver dut a terme les simulacions de dinàmica molecular per a cada una de les variants de l'enzim obtingudes, es representen les gràfiques de distància catalítica i de RMSD de la base de Schiff. A més a més, es proporcionen les imatges obtingudes a través de PyMol de cada una de les rèpliques. Aquestes facilitaran la comprensió les gràfiques. Es relacionarà la RMSD de la base de Schiff amb la seva superposició mostrada en les imatges. A més a més, es podran visualitzar fàcilment les distàncies catalítiques i les diferències entre elles. Aquestes informacions es recullen a continuació.

Per a cada variant, s'analitzen i es descriuen cada un dels comportaments que mostra qualsevol de les rèpliques de cada variant. Basant l'anàlisi en les imatges i gràfiques resultants per a cada una de les variants de la RA95.5-8F, doncs, s'utilitza el mateix ordre seguit a l'apartat anterior. A més a més, s'adjunten els valors de la mitjana de la distància catalítica per a cada rèplica, fet que ajudarà a comprendre el comportament de cada una d'elles i, així mateix, la visió general.

Es consideraran actives, per tant, que presenten una conformació òptima per a la catàlisi, les rèpliques amb una distància catalítica d'aproximadament $4 \text{ \AA}^{[5]}$.

4.1. “Variant 75”: RA95.5-8F amb la posició 75 revertida (P75R)

Aquesta variant presenta revertida la posició 75, és a dir, conté una arginina en comptes d’una prolina. Aquesta mutació es troba desplaçada +2 posicions en seqüència del SPM. Si s’analiza la gràfica de distància catalítica de la variant 75 (veure *Figura 7*), es pot observar que les rèpliques 0, 1 i 2 mantenen una distància fixa d’aproximadament 2.9 Å durant la major part de la simulació. Per altra banda, les dues restants (rèpliques 3 i 4) presenten grans canvis en la distància catalítica, que arriba a assolir fins a 8 Å. Això farà que les seves respectives distàncies catalítiques mitjanes siguin més elevades. Pel que fa al valor de RMSD de la base de Schiff d’aquestes dues rèpliques, es presenta molt variant durant el temps de simulació. Aquest fet indica que no hi haurà massa superposició i que els seus valors de desviació estàndard seran majors. Així mateix, i d’una forma més visual, es poden explicar els comportaments de les rèpliques mitjançant les imatges obtingudes amb PyMol, relacionar-les amb les dues gràfiques descrites. Aquestes rèpliques 3 i 4 són les que presenten més variacions tant de distància catalítica com de RMSD degut a una major flexibilitat de la base de Schiff. D’aquestes dues rèpliques, la 4 adopta valors més elevats, el que representa que els àtoms de la base de Schiff ocupen posicions més diferents a les de referència. Respecte a les tres primeres rèpliques, totes tenen un comportament més constant (desviació estàndard menor). Tot i això, presenten algunes diferències entre elles. La rèplica 0 és la que presenta valors més baixos. Mostra algunes fluctuacions a l’inici de la simulació, però després adopta el comportament més constant de les cinc rèpliques (desviació estàndard menor). Per aquest motiu és la que presenta la major superposició i, a més a més, un menor desplaçament respecte els residus de referència. La RMSD de la rèplica 1 es manté constant durant la simulació, però no tant com en el tram de la rèplica 0. Degut a això, la base de Schiff d’aquesta rèplica també presenta una bona superposició de les estructures visitades al llarg del temps de simulació, però no tant com a la rèplica 0. Finalment, la rèplica 2 mostra un interval variant durant els primers 400 ns. Així doncs, es pot explicar el valor de les distància catalítica mitjana i, sobretot de les desviacions estàndard. En les imatges de PyMol, la base de Schiff adopta dues conformacions diferents. Una equival als primers 400 ns de simulació (valor menor de RMSD) i l’altre equival a la resta de la simulació. Tots els comportaments de cada rèplica es poden apreciar clarament a partir de la *Taula 3*, que recull la mitjana de la distància catalítica per cada rèplica i la corresponent desviació estàndard.

Taula 3. Mitjana de distància catalítica i desviació estàndard per cada rèplica

Variant 75	<i>Rèplica 0</i>	<i>Rèplica 1</i>	<i>Rèplica 2</i>	<i>Rèplica 3</i>	<i>Rèplica 4</i>
Distància (Å)	2.64 ± 0.26	2.75 ± 0.62	2.97 ± 0.90	4.02 ± 1.22	4.81 ± 1.56

Així doncs, es pot concloure que la variant 75 mostra dos comportaments clars. Les tres primeres rèpliques presenten una conformació del centre actiu òptim per a la catàlisi (són actives), mentre que les dues darreres no són tant competents des d’un punt de vista de la catàlisi. Tot i així, en 3 de 5 rèpliques la base està ben posicionada per la catàlisi.

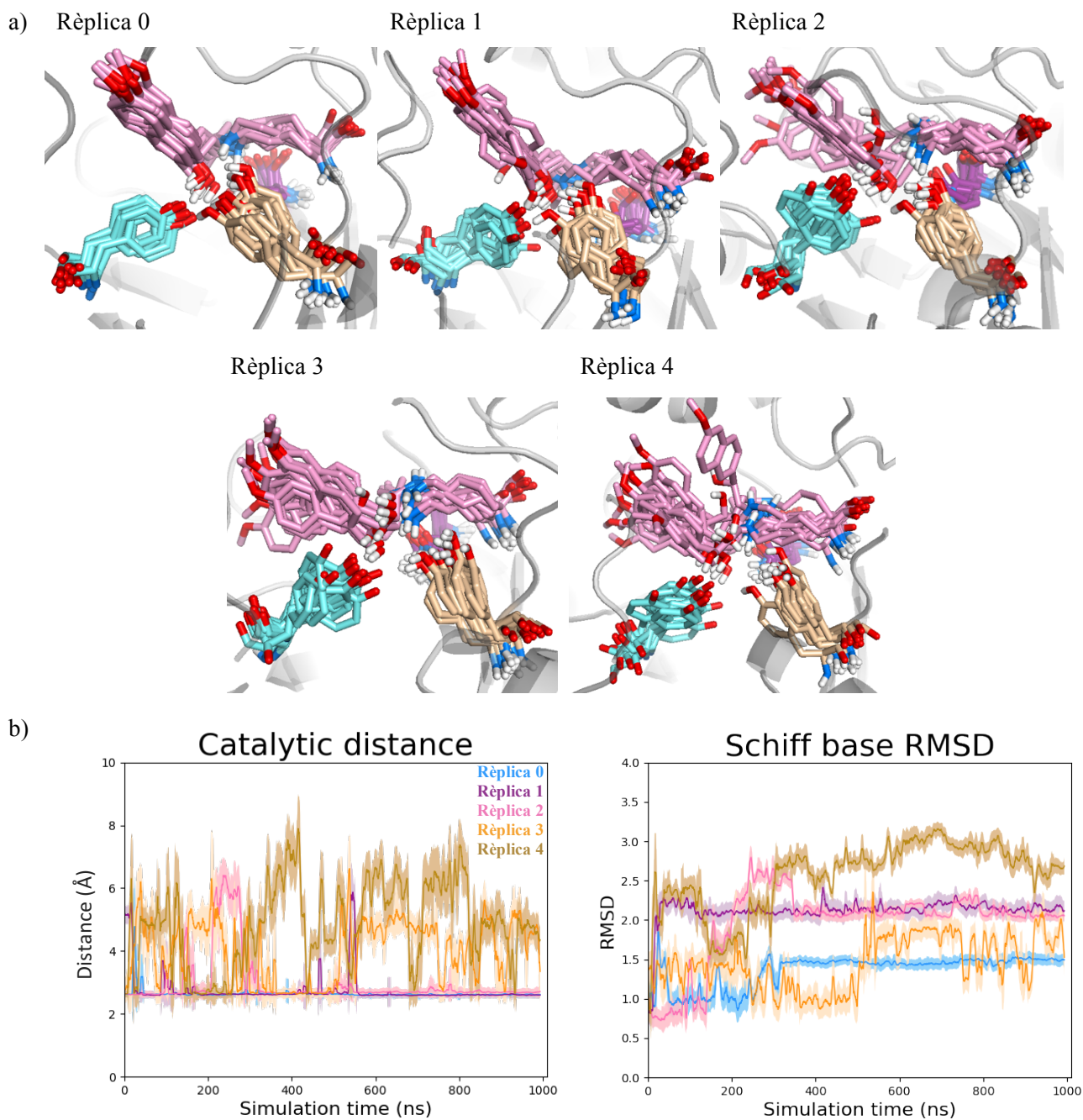


Figura 7. (a) Representació per superposició de la trajectòria de la tètrada catalítica per a cada rèplica de la RA95.5-8F amb la posició 75 revertida (P75R). En beix la Tyr50, en rosa la base de Schiff en la posició 82 i en cian la Tyr179. En lila, la Asn109, en segon pla per finalitats de visualització. (b) Gràfiques de la distància catalítica en Àngstroms (Å) i la RMSD (Å) de la base de Schiff. La rèplica 0 està representada en blau; la rèplica 1 en lila; la rèplica 2 en rosa; la rèplica 3 en taronja i la rèplica 4 en marró.

4.2. “Variant 90”: RA95.5-8F amb la posició 90 revertida (D90N)

Aquesta variant presenta revertida la posició 90, és a dir, conté una asparagina en comptes d’un àcid aspàrtic. Aquesta mutació també es troba desplaçada +6 posicions en seqüència del SPM. Observant la gràfica de distància catalítica de la variant 90 (veure *Figura 8*), es veu a simple vista com de lluny estan la base Tyr180 i la base de Schiff de la rèplica 3. La seva distància es manté constant a aproximadament 5 Å durant tot el temps de simulació, però no és gens similar a la resta de rèpliques. Això és degut al posicionament del β -alcohol a l’altra cara de la base de Schiff. Pel que fa les altres rèpliques, mostren aquesta distància també constant, però partint d’aproximadament 2.9 Å durant els 1000 ns. Cal notificar que aquest valor constant de distància és el mateix que el que mostraven les rèpliques corresponents de la variant 75. Per una banda, les rèpliques 0 i 2, mantenen el valor de distància catalítica durant els primers 400ns i, posteriorment, va variant. Arriben a assolir una distància de fins a 8 Å, donat que la base Tyr180 presenta una major mobilitat. Per altra banda, les rèpliques 1 i 4 també tenen un comportament similar entre elles. Es pot considerar que mantenen la distància catalítica al llarg de la simulació. Per aquest motiu tenen els valors de distància catalítica mitjana més pròxims als 2.9 Å. Per la gràfica de RMSD de la base de Schiff, la rèplica 3 és la que presenta un valor constant més elevat i diferent al de les altres rèpliques. Observant la imatge de PyMol corresponent a la rèplica, s’explica el comportament de les gràfiques anteriorment descrites. Es pot explicar d’una forma visual el valor tant elevat de RMSD d’aquesta rèplica, ja que la seva base de Schiff és la que ocupa una posició més diferent a la de referència (l’hidroxil està molt lluny de la base). A més a més, és la que presenta un valor més constant (desviació estàndard menor) i, per tant, la que mostra més superposició. Les rèpliques 1 i 4 i, així mateix, la 0 i la 2, tenen un comportament en la gràfica de RMSD respectivament semblant. La primera parella de rèpliques té majors valors de RMSD que la segona, però també es mantenen lleugerament més constants. El valor de RMSD de la base de Schiff de la rèplica 0 durant la simulació és més constant que el de la rèplica 2, per aquest motiu aquesta segona no presenta tanta superposició, com es pot veure a la imatge. Pel que fa a les rèpliques 1 i 4 són les que mantenen més la distància catalítica durant els 1000 ns. Ara bé, com es pot veure a les imatges, la rèplica 1 presenta més superposició que la rèplica 4. Això ajuda a explicar que les seves desviacions estàndards siguin mínimament menors. Mitjançant la *Taula 4* es poden entendre tots els comportaments de cada rèplica.

Taula 4. Mitjana de distància catalítica i desviació estàndard per cada rèplica

Variant 90	<i>Rèplica 0</i>	<i>Rèplica 1</i>	<i>Rèplica 2</i>	<i>Rèplica 3</i>	<i>Rèplica 4</i>
Distància (Å)	3.83 ± 1.61	3.04 ± 1.09	5.02 ± 1.95	5.24 ± 0.39	2.87 ± 0.73

D’aquesta manera es pot concloure que les rèpliques 1 i 4 de la variant 90 són actives ja que presenten distàncies catalítiques al voltant de 3 Å. Les rèpliques 0 i 2 presenten els residus catalítics i la base de Schiff ben posicionats per la catàlisi només durant els primers 400ns del temps de simulació. La rèplica 3 és inactiva degut al posicionament de l’hidroxil de la base de Schiff lluny de la base Tyr180 .

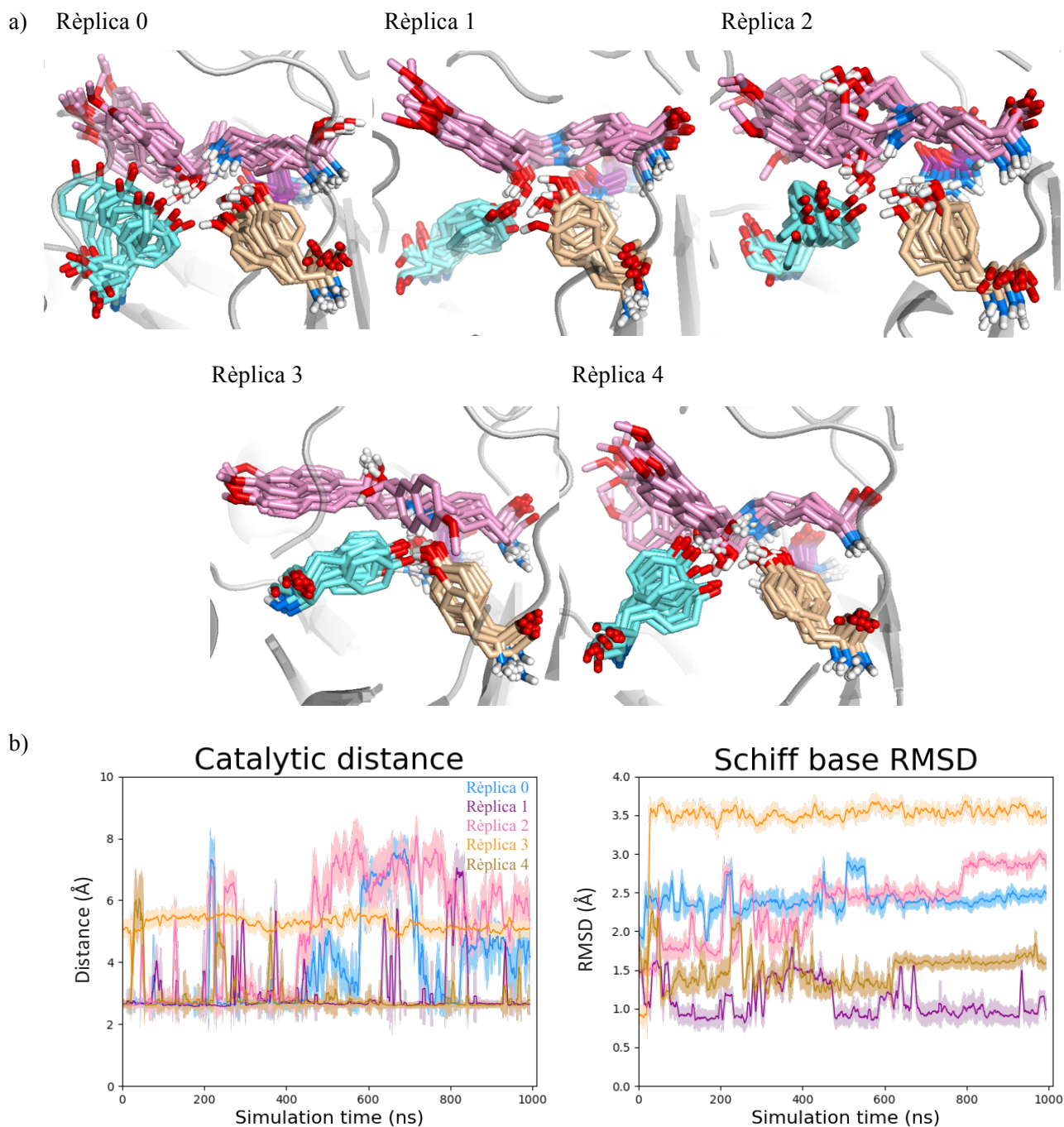


Figura 8. (a) Representació per superposició de la trajectòria de la tètrada catalítica per a cada rèplica de la RA95.5-8F amb la posició 90 revertida (D90N). En beix la Tyr50, en rosa la base de Schiff en la posició 82 i en cian la Tyr179. En lila, la Asn109, en segon pla per finalitats de visualització. (b) Gràfiques de la distància catalítica en Àngstroms (Å) i la RMSD (Å) de la base de Schiff. La rèplica 0 està representada en blau; la rèplica 1 en lila; la rèplica 2 en rosa; la rèplica 3 en taronja i la rèplica 4 en marró.

4.3. “Variant 135”: RA95.5-8F amb la posició 135 revertida (E135N)

Aquesta variant presenta revertida la posició 135, és a dir, conté una asparagina en comptes d’un àcid glutàmic. Aquesta mutació es troba desplaçada +3 posicions en seqüència del SPM. La variant 135 presenta dues rèpliques (2 i 4) amb una valor de distància catalítica que es pot considerar constant al llarg de la simulació (veure *Figura 9*). Cal notificar, altra vegada, que la distància base torna a ser d’aproximadament 2.9 Å. Tot i això, la rèplica 4 comença a presentar petites fluctuacions a partir dels 700 ns. Degut això, les dues rèpliques en qüestió mostren els menors valors de distància catalítica mitjana i, a més a més molt similars, però la de la rèplica 4 és lleugerament major que la de la rèplica 2. La resta de rèpliques presenten més variacions durant la simulació, essent la rèplica 0 la que mostra més fluctuacions. A més a més, és la que adopta valors de distància catalítica més elevats durant més temps donat que la Tyr180 presenta una major flexibilitat i la base de Schiff pot adoptar dues conformacions diferents. És per aquest motiu que presenta el valor de distància catalítica mitjana major (veure Taula 5). Les rèpliques 1 i 3 tenen un comportament similar pel que fa a la seva distància catalítica al llarg del temps, tot i que la base Tyr180 presenta certa flexibilitat. En referència a la gràfica de RMSD de la base de Schiff, la rèplica 2 és la que mostra els menors valors durant el temps, fet que significa que és la que adopta una posició més similar a la de referència i, a més a més, la més estable durant la simulació. Aquesta rèplica, juntament amb la rèplica 1, són les que tenen aquest valor més constant al llarg del temps i, per aquest motiu, són les rèpliques que presenten unes desviacions estàndards menors. Utilitzant les imatges obtingues amb PyMol es pot veure com aquestes dues rèpliques són les que presenten una major superposició de la base de Schiff, encara que la rèplica 2 en presenti una mica més que la rèplica 1, degut al caràcter més constant de RMSD. Les altres tres rèpliques mostren fluctuacions clares en gràfica de RMSD de la base de Schiff, per tant aquesta adopta varies conformacions al llarg de les simulacions, com es pot veure a les corresponents imatges. Les rèpliques 0 i 3 són les que mostren els valors més elevats de RMSD, tot i que els de la rèplica 0 són lleugerament majors i menys constants (desviació estàndard major). Això significa, també, que la base de Schiff d’aquesta rèplica és la que presenta una posició més diferent a la de referència i per tant, una major flexibilitat (menys superposició). Les rèpliques 1 i 4 són similars pel que fa al valor de RMSD i la seva constància al llarg del temps, per això les superposicions d’estructures són també similars. La rèplica 2 és la que presenta més superposició de totes les rèpliques.

Taula 5. Mitjana de distància catalítica i desviació estàndard per cada rèplica

<i>Variant 135</i>	<i>Rèplica 0</i>	<i>Rèplica 1</i>	<i>Rèplica 2</i>	<i>Rèplica 3</i>	<i>Rèplica 4</i>
<i>Distància (Å)</i>	5.00 ± 1.52	4.20 ± 0.86	2.98 ± 0.76	4.40 ± 1.06	3.17 ± 1.10

Així doncs, es pot concloure que per la variant 135 hi ha dues variant actives (2 i 4). Les tres variants restants presenten una major desviació de la distància catalítica i, per tant, són menys competents des d’un punt de vista catalític.

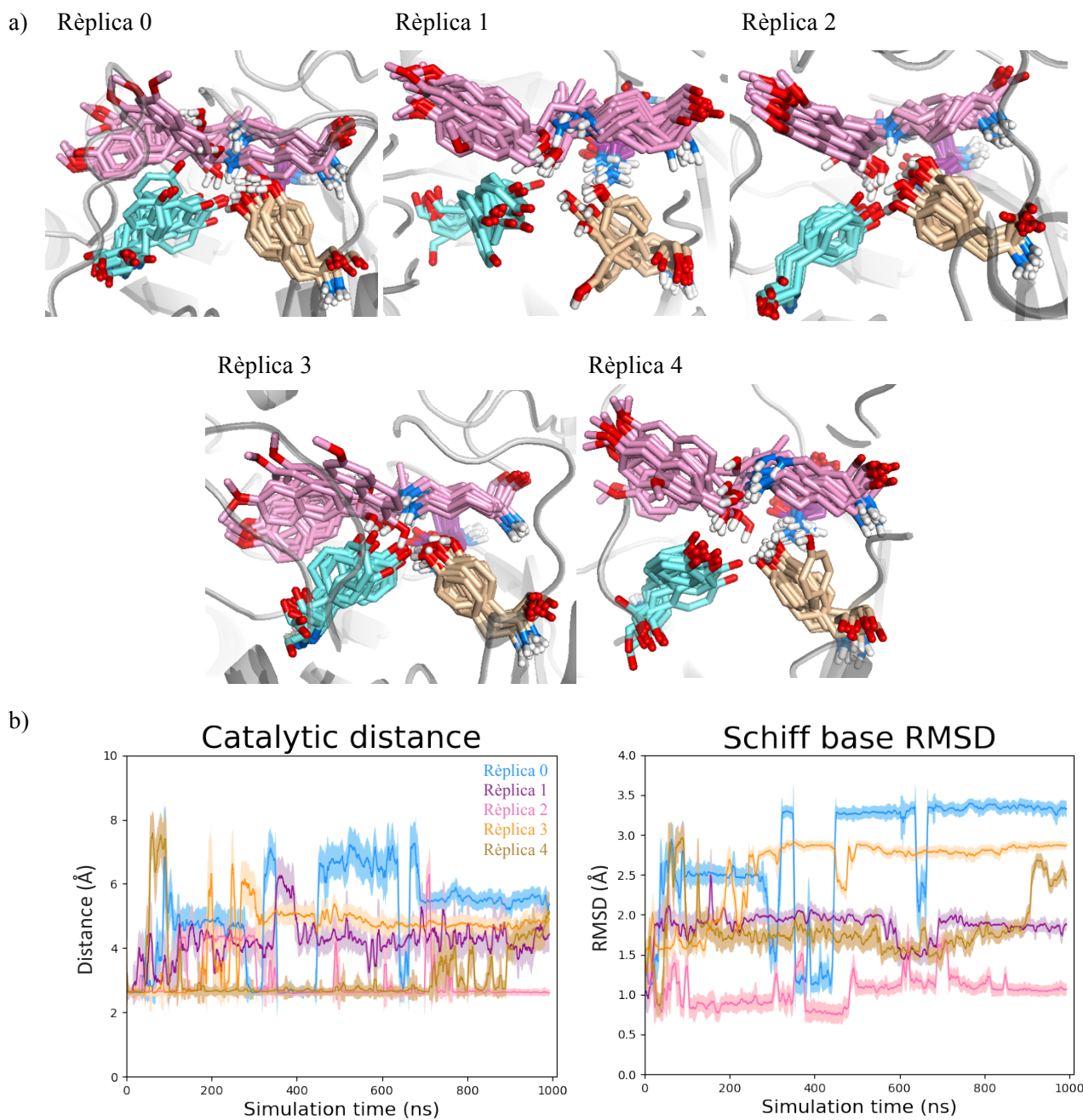


Figura 9. (a) Representació per superposició de la trajectòria de la tètrada catalítica per a cada rèplica de la RA95.5-8F amb la posició 135 revertida (E135N). En beix la Tyr50, en rosa la base de Schiff en la posició 82 i en cian la Tyr179. En lila, la Asn109, en segon pla per finalitats de visualització. (b) Gràfiques de la distància catalítica en Àngstroms (Å) i la RMSD (Å) de la base de Schiff. La rèplica 0 està representada en blau; la rèplica 1 en lila; la rèplica 2 en rosa; la rèplica 3 en taronja i la rèplica 4 en marró.

4.4. “Variant 151”: RA95.5-8F amb la posició 151 revertida (G151S)

Aquesta variant presenta revertida la posició 151, és a dir, conté una serina en comptes d’una glicina. Aquesta mutació també es troba desplaçada +6 posicions en seqüència del SPM. Pel que fa a la gràfica de distància catalítica es poden apreciar tres comportaments diferents (veure *Figura 10*). Per una banda, la rèplica 2 es mostra poc variant al llarg del temps de simulació, però amb valors d’aproximadament 4 Å. Això és degut al posicionament relativament llunyà del β-alcohol i la base de Schiff, degut majoritàriament a la mobilitat de la base Tyr180. La resta de rèpliques presenten una distància molt més constant durant els 1000 ns. Cal esmentar altra vegada que la distància base torna a ser d’aproximadament 2.9 Å. D’aquestes quatre rèpliques, les rèpliques 1 i 4 mantenen la distància mitjanament estable però amb fluctuacions puntuals de fins a 8 Å que provenen d’una segona conformació que adopta la base de Schiff. Això fa augmentar considerablement el seu valor de la mitjana de distància catalítica (veure Taula 6). Per altra banda, les rèpliques 0 i 3, no mostren tantes variacions. Per aquest motiu, aquestes dues últimes rèpliques tenen una distància catalítica mitjana menor que la de les rèpliques 1 i 4. Pel que fa a la gràfica de RMSD de la base de Schiff, totes les rèpliques mostren intervals constants que varien durant certs trams del temps de simulació. Es poden entendre les gràfiques de RMSD de la base de Schiff i distància catalítica d’una manera més visual mitjançant les imatges obtingudes amb PyMol. Les rèpliques 1 i 4 són les que presenten els valors de RMSD menys constants (desviacions estàndards majors) i menys superposició. Tot i això, la rèplica 4 mostra valors molt majors, indicant que la seva base de Schiff adopta una posició més diferent respecte la de referència. Les rèpliques restants mostren valors molt més constant de RMSD. La rèplica 0 mostra un interval molt constant, des dels 300ns al final de la simulació, amb un valor de RMSD d’aproximadament 2 Å. Ara bé, també presenta un segon tram menys constant però d’un valor menor. Així doncs, veient la imatge corresponent a aquesta rèplica, es poden veure les dues conformacions que adopta la base de Schiff, que corresponen als dos intervals de la simulació. La primera, de RMSD menor, serà més similar a la de referència; mentre que la segona, de RMSD major, diferirà més de la posició de referència. El mateix efecte passa per la rèplica 1. L’interval de temps entre 100 i 300ns equival a una conformació diferent a la que adopta la resta de temps. Les rèpliques 3 i 4 també presenten dues conformacions diferents de la base de Schiff que equivalen a un canvi de valor de la RMSD. La RMSD de les rèpliques 2 i 3 tenen comportaments molt similars, per això també tenen valors de desviació estàndards i superposició de la base de Schiff semblants.

Taula 6. Mitjana de distància catalítica i desviació estàndard per cada rèplica

Variant 151	Rèplica 0	Rèplica 1	Rèplica 2	Rèplica 3	Rèplica 4
Distància (Å)	2.74 ± 0.55	3.42 ± 1.64	3.63 ± 0.86	3.05 ± 1.09	3.90 ± 1.69

Així doncs, les rèpliques 0 i 3 de la variant 151 presenten una distància catalítica al voltant de 3 Å (actives). Les rèpliques 1 i 4 presenten els residus catalítics i la base de Schiff ben posicionats per la catàlisi temporalment. A causa de les grans distàncies catalítiques que adopta la rèplica 4 en gran part de la simulació, es considera que el seu centre catalític no està massa ben posicionat per la catàlisi. La rèplica 2 es considera inactiva.

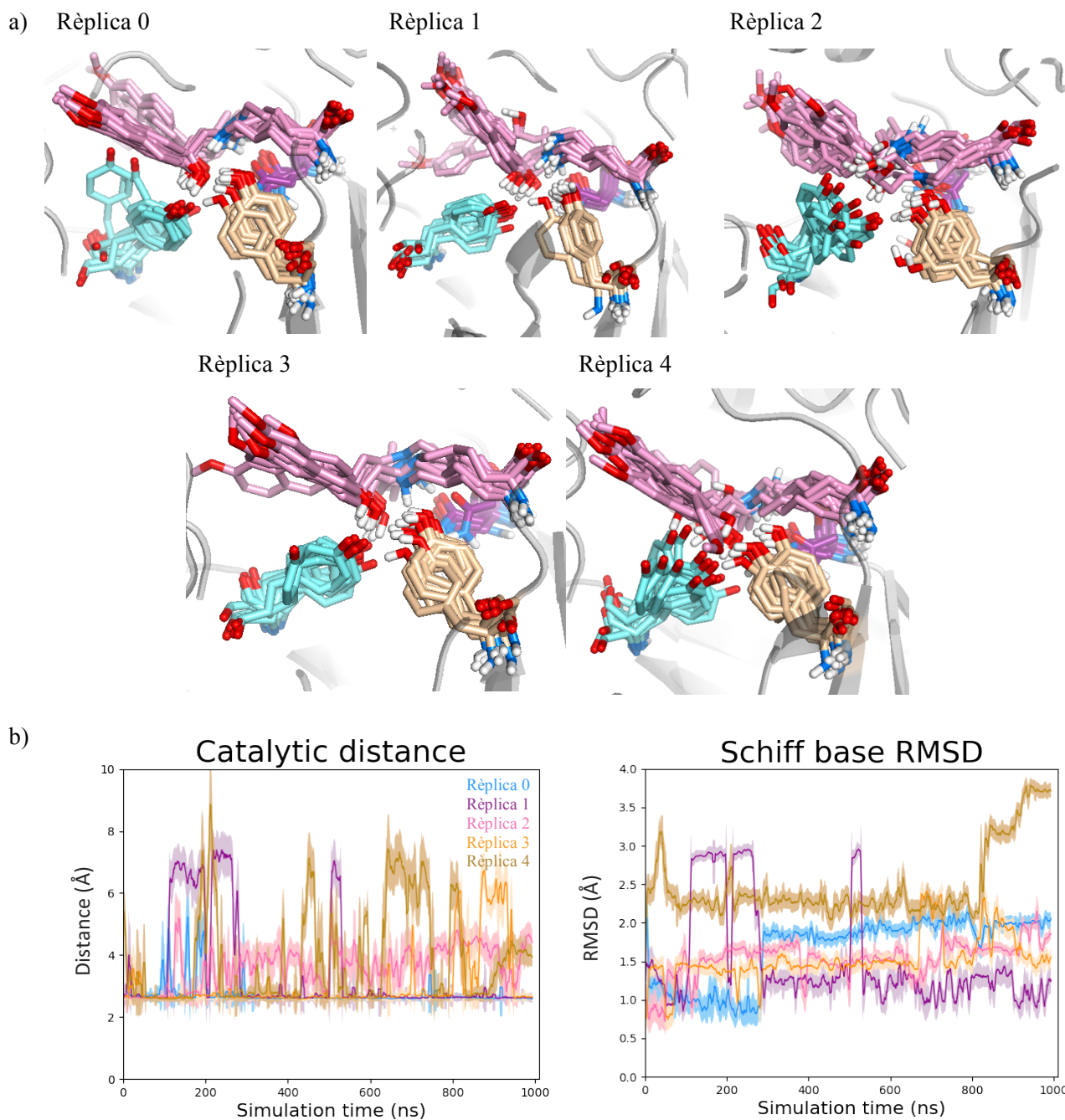


Figura 10. (a) Representació per superposició de la trajectòria de la tètrada catalítica per a cada rèplica de la RA95.5-8F amb la posició 151 revertida (G151S). En beix la Tyr50, en rosa la base de Schiff en la posició 82 i en cian la Tyr179. En lila, la Asn109, en segon pla per finalitats de visualització. (b) Gràfiques de la distància catalítica en Àngstroms (Å) i la RMSD (Å) de la base de Schiff. La rèplica 0 està representada en blau; la rèplica 1 en lila; la rèplica 2 en rosa; la rèplica 3 en taronja i la rèplica 4 en marró.

4.5. “Variant 75_135”: RA95.5-8F amb les posicions 75 i 135 revertides (P75R, E135N)

Aquesta variant presenta revertides les posicions 75 i 135, és a dir, conté una asparagina i una arginina en comptes d’un àcid glutàmic i una prolina. Aquestes dues mutacions es troben desplaçades a +2 i +3 posicions en seqüència del SPM. Aquesta és la primera variant mostrada que prové de revertir dues posicions. Observant la gràfica de distància catalítica es pot veure un comportament diferent comparat amb els anteriors casos esmentats (veure *Figura 11*). Tant sols a la rèplica 4 se li pot intuir una certa constància a la gràfica de distància catalítica durant el temps de simulació. Per aquest motiu, és també la única rèplica que presenta un valor mitjà que es pot considerar baix, pròxim als 3 Å. Cal tornar a recalcar que la distància base torna a ser de 2.9 Å. En qualsevol de les altres rèpliques hi ha fluctuacions durant el transcurs de la simulació, degut a la mobilitat de la base Tyr180. La rèpliques 1, 2 i 3 presenten valors majors, per això les seves distàncies catalítiques mitjanes són també majors. La rèplica 0, encara que també presenta un comportament fluctuant, assoleix valors menors, fent que la seva mitjana sigui també menor. Pel que fa a la gràfica de RMSD de la base de Schiff, totes presenten valors lleugerament alts en comparació a les altres variants descrites. Això significa que la seva base de Schiff adopta posicions diferents respecte a la de referència. A més a més, hi abunden les fluctuacions del valor de RMSD al llarg de la simulació, que correspondrà a valors de desviació estàndard relativament elevats i per tant, la base de Schiff esdevé més flexible. La rèplica 1 presenta un interval molt constant de RMSD entre els 200 fins als 600 ns. És per aquest motiu que en la imatge corresponent s’hi pot observar certa superposició. Tot i això la rèplica 4 és la que mostra valors més constants pel que fa a la totalitat de la simulació (desviació estàndard menor) i, a més a més, amb els valor més baixos. Això significa que, de les quatre rèpliques, és la que adopta la conformació de la base de Schiff més similar a la de referència. Encara que també sigui la rèplica més constant, no presenta massa superposició. La rèplica 0 és la que presenta més variació de RMSD i valors més alts. Per això és la que té el valor més elevat de desviació estàndard i un major flexibilitat de la base de Schiff (menys superposició). Les rèpliques 2 i 3 presenten un comportament de RMSD molt similar, tant pel que fa als valors com per la seva constància. Això explica la seva similitud de valor estàndard i així mateix, les superposicions de les estructures.

Taula 7. Mitjana de distància catalítica i desviació estàndard per cada rèplica

Variant 75_135	<i>Rèplica 0</i>	<i>Rèplica 1</i>	<i>Rèplica 2</i>	<i>Rèplica 3</i>	<i>Rèplica 4</i>
Distància (Å)	4.50 ± 1.46	6.55 ± 1.36	5.49 ± 1.43	5.58 ± 1.45	3.30 ± 1.13

En el cas de la variant 75_135, la única rèplica que es pot considerar que presenta una conformació del centre actiu òptima per la catàlisi és la 4. Les rèpliques restants són no són competents des del punt de vista catalític ja que la base de Schiff presenta gran mobilitat. Així doncs, en només 1 de 5 rèpliques la base està ben posicionada per la catàlisi.

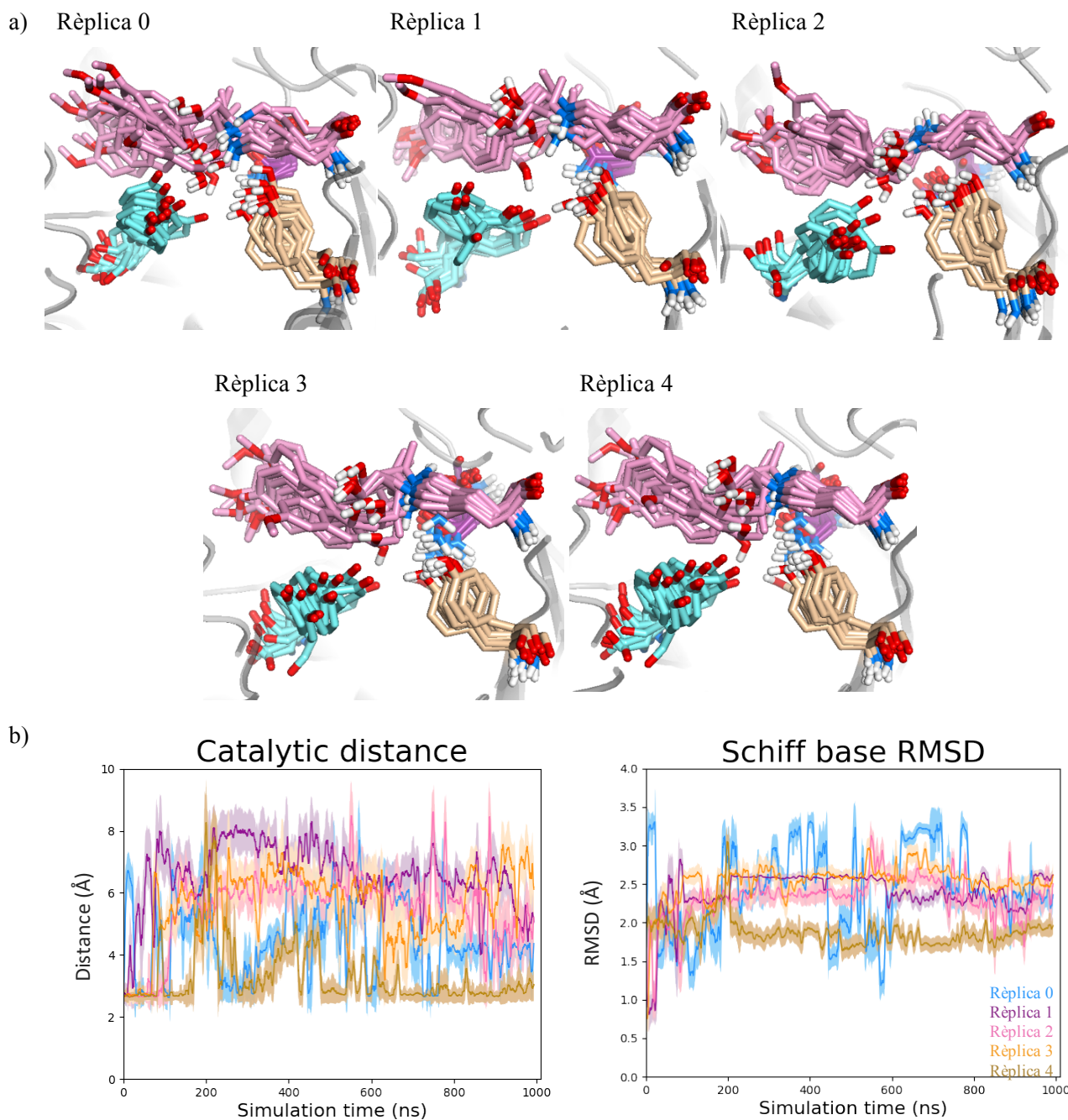


Figura 11. (a) Representació per superposició de la trajectòria de la tètrada catalítica per a cada rèplica de la RA95.5-8F amb les posicions 75 i 135 revertides (P75R, E135N). En beix la Tyr50, en rosa la base de Schiff en la posició 82 i en cian la Tyr179. En lila, la Asn109, en segon pla per finalitats de visualització. (b) Gràfiques de la distància catalítica en Àngstroms (Å) i la RMSD (Å) de la base de Schiff. La rèplica 0 està representada en blau; la rèplica 1 en lila; la rèplica 2 en rosa; la rèplica 3 en taronja i la rèplica 4 en marró.

4.6. “Variant 90_151”: RA95.5-8F amb les posicions 90 i 151 revertides (D90N, G151S)

Aquesta variant presenta revertides les posicions 90 i 151, és a dir, conté una asparagina i una serina en comptes d'un àcid aspàrtic i una glicina. Aquestes mutacions es troben desplaçades +6 posicions en seqüència del SPM. Tot i que per aquesta rèplica també es reverteixen dues posicions, el comportament és totalment diferent a la variant anterior. Observant la gràfica de distància catalítica es pot veure que les rèpliques 2, 3 i 4 tenen un comportament molt constant (veure *Figura 12*). D'aquestes, la rèplica 2 presenta un període fluctuant a partir dels 100 fins als 300 ns, amb valors de distància catalítica alts, arribant a assolir els 8 Å. Aquests valors alts venen donats per la mobilitat de la Tyr180. La rèplica 4 presenta més variacions que la 2, però amb valors menys alts, tant sols arribant a assolir els 6 Å de distància catalítica. La seva base Tyr180 no presenta moviments tant bruscs com els de la rèplica 2. Els respectius valors de distància catalítica mitjana d'aquestes rèpliques són molt similars entre ells. La rèplica 3 és la que presenta la distància catalítica més constant durant la simulació. És per aquest motiu que la seva mitjana és la més similar als 2.9 Å de la distància base. Pel que fa a la resta de rèpliques, la 0 adopta valors molt elevats i fluctuants de distància catalítica mentre que la rèplica 1, també pateix moltes variacions de distància catalítica però de valor molt menor. Pel que fa la gràfica de RMSD de la base de Schiff, la rèplica 1 manté un valor alt i variant durant el temps de simulació (desviació estàndard major), fet que indica que adopta una posició molt diferent a la de referència i, per tant, la seva base de Schiff és molt flexible. Les altres rèpliques tampoc mostren la RMSD constant en cap moment, però sí que mantenen valors menors. Així doncs, no s'hauria d'esperar superposició de cap de les rèpliques. Ara bé, observant les imatges obtingudes amb PyMol de les rèpliques 2 i 3, es pot intuir certa superposició. Realment n'hi tanta com a la resta de rèpliques. Degut a finalitats de visualització de la distància catalítica, la figura no permet veure el ventall que forma la base de Schiff a la part esquerra de les imatges corresponents, deixant intuir una falsa superposició. Així doncs, cap de les cinc rèpliques presenta superposició en la seva base de Schiff, ni tant sols les rèpliques 2 i 3. Tornant a la gràfica de RMSD de la base de Schiff, no es pot apreciar quina de les rèpliques és la més constant. Ara bé, veient la *Taula 8*, es pot intuir que les rèpliques amb una desviació estàndard menor són, també, les de RMSD més constant. Aquestes són la rèpliques 1 i 3. La rèplica 2 és la que mostra els valors de RMSD menors, per tant la seva base de Schiff ocuparà un posició més similar a la de referència.

Taula 8. Mitjana de distància catalítica i desviació estàndard per cada rèplica

Variant 90_151	<i>Rèplica 0</i>	<i>Rèplica 1</i>	<i>Rèplica 2</i>	<i>Rèplica 3</i>	<i>Rèplica 4</i>
Distància (Å)	6.04 ± 1.66	3.65 ± 0.79	3.13 ± 1.30	2.84 ± 0.48	3.17 ± 0.93

Així doncs, es pot concloure que per la variant 90_151, les rèpliques 2, 3 i 4 són totalment actives degut a la seva distància catalítica similar als 3 Å. Les dues primeres no presenten els residus catalítics i la base de Schiff ben posicionats per la catàlisi.

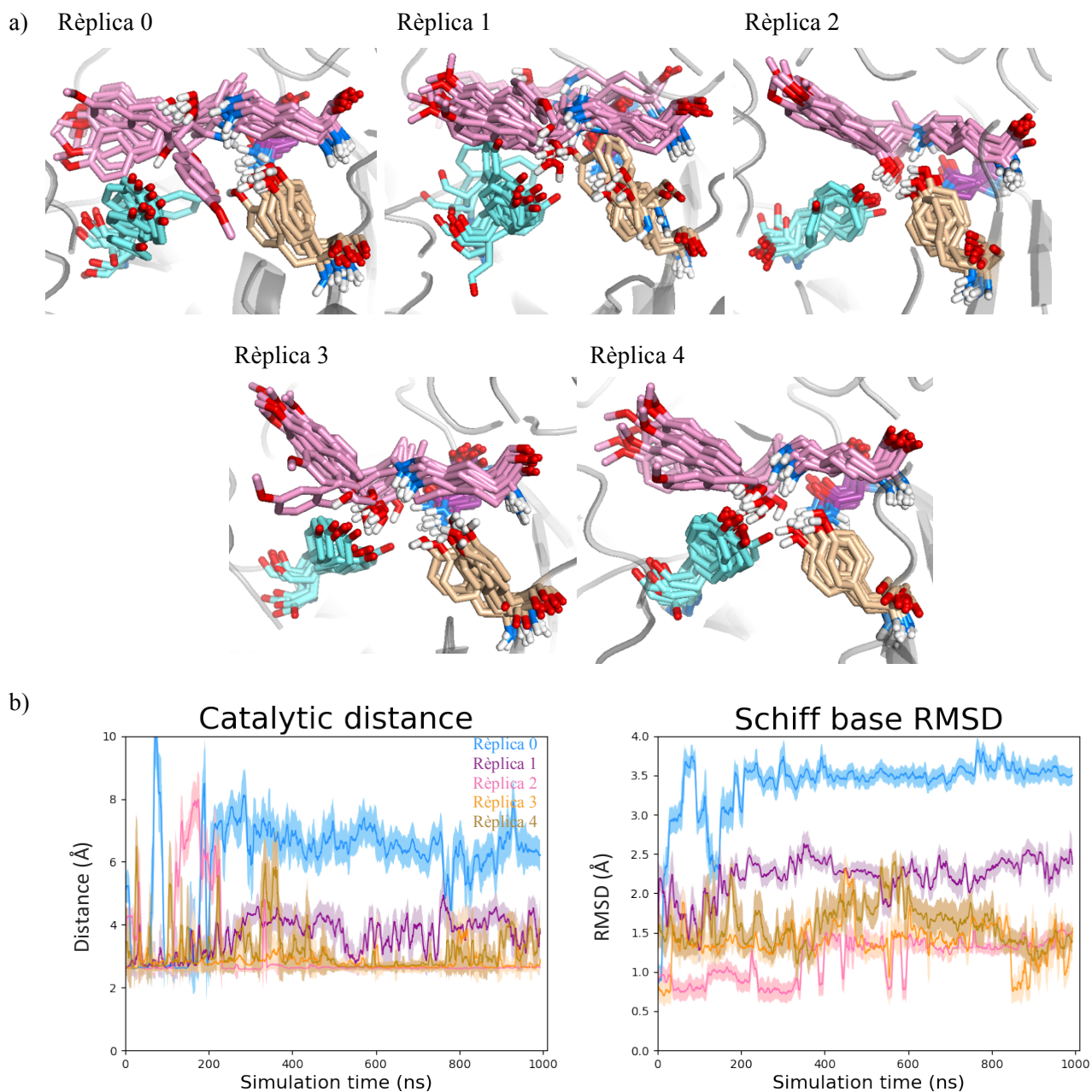


Figura 12. (a) Representació per superposició de la trajectòria de la tètrada catalítica per a cada rèplica de la RA95.5-8F amb les posicions 90 i 151 revertides (D90N, G151S). En beix la Tyr50, en rosa la base de Schiff en la posició 82 i en cian la Tyr179. En lila, la Asn109, en segon pla per finalitats de visualització. (b) Gràfiques de la distància catalítica en Àngstroms (Å) i la RMSD (Å) de la base de Schiff. La rèplica 0 està representada en blau; la rèplica 1 en lila; la rèplica 2 en rosa; la rèplica 3 en taronja i la rèplica 4 en marró.

4.7. Comportament general

Anteriorment s'han descrit i detallat cada una de les imatges de PyMol i les gràfiques de distància catalítica i RMSD de la base de Schiff de cada una de les variants. Tota aquesta informació ha permès predir la seva activitat catalítica. Tot i això, per entendre el comportament de cada una de les variants descrites anteriorment cal recordar a quantes posicions respecte del SPM es troba cada un dels residus revertits. De més a menys propers: Arg75 (+2), Asn135 (+3), Asn90 (+6) i Ser151 (+6).

Començant per les variants a les quals se'ls ha revertit els residus amb les posicions més llunyanes al SPM (+6), concretament les variants 90 i 151, tant sols presenten una rèplica totalment inactiva cadascuna (1 de 5). Així doncs es podria dir que les dues variants són molt actives en absència d'aquestes mutacions. Per la variant 135, que se li ha revertit el residu que es troba a +3 posicions respecte el SPM, mostra tres rèpliques completament inactives (3 de 5), indicant que aquesta posició té un efecte més rellevant que les posicions 90 i 151. De totes les rèpliques de la variant 75, tant sols dues són totalment inactives (2 de 5). Aquesta variant se li havia revertit un residu situat a +2 posicions del SPM. Aquestes dues variants són menys actives que les dues anteriors, les quals se'ls ha revertit una mutació situada a +6 posicions del SPM.

Es pot començar a entreveure, doncs, que el número de posicions respecte el SPM que es troba el residu revertit podria estar relacionat amb l'activitat catalítica de l'enzim en qüestió. A més a més, no es poden detectar grans canvis estructurals a la tètrada catalítica per les variants descrites que tenen mutacions desplaçades unes 6 posicions del camí predit per SPM (variants 90 i 151). Sí és cert que els residus catalítics poden adoptar conformacions addicionals al llarg de la simulació, sempre situant la base de Schiff a la base catalítica. A continuació, es segueix analitzant les variants que contenen més d'una reversió.

Per la variant 90_151, tres de les rèpliques es presenten actives. Per altra banda, la variant 75_135, només mostra una de les rèpliques actives. Aquesta diferència d'activitat es pot explicar per la diferència de posicions respecte el SPM. Pel primer cas, les dues posicions que es reverteixen es troben a +6 del SPM i per tant, es pot suposar que la variació de l'activitat catalítica serà molt menor que si es revertissin posicions més properes. Aquest és el cas de la variant 75_135. Els residus que es reverteixen estan situats a +2 i +3 posicions respecte el SPM. El fet que només una de les rèpliques sigui activa indica que l'impacte d'aquestes mutacions sobre l'activitat catalítica és molt major que pel cas de la variant 90_151. Pel que fa a la tètrada catalítica d'aquestes variants, tampoc ressenten grans canvis estructurals, però sí és cert que mostren menys superposició que en les variants anteriors, indicant una major flexibilitat de la base de Schiff i la base de Tyr180.

5. *Conclusions*

In this project, the effect of four residues mutated of the enzyme retro-aldolase 95 (RA95) has been analysed. First, the PyMol program has been used to reverse the mutation on those studied residues. Second, a computational protocol has been used to carry out a molecular dynamics (MD) simulation for each variant obtained. Finally, two different graphics for the analysis have been obtained: the catalytic distance and Schiff base fluctuation as computed with the RMSD. For a better visualization, the PyMol program has been used again.

The main conclusions obtained from all the MD simulations and their analysis can be summarized as follows:

- The catalytic distance is closely related to the number of positions far from the SPM where the residue mutated is situated. It has been concluded that the effect of the mutations on the catalytic distance is lower when the number of positions far from the SPM is higher.
- The variants that have only one residue mutated show a better overlay of the sampled conformation for the Schiff base than in the other doubly mutated variants. Thus, this deviation (and the large Schiff base RMSD) is related to the number of mutations applied on the enzyme. A higher flexibility and deviation from the initial positioning have a dramatic impact into the catalytic activity of the enzyme.
- The rest of the catalytic tetrad residues (Asn110 and Tyr51) do not show important structural changes after the application of the mutated residues in any of the MD simulations performed and analysed.
- The SPM is a great method to determine which positions of the enzyme are relevant for the retro-aldol activity.

The catalytic distance is an important factor to take into account for the catalytic activity of each enzyme. Despite of that, there are also other factors such as stability or flexibility. To further validate our predictions and know exactly the influence of the mutations in the catalytic activity, experimental test should be done.

Altogether our MD analysis has shown that the main working hypothesis of this project “the residues situated far from the SPM have a minor effect on the enzyme retro-aldol activity compared with those situated in adjacent positions” is valid.

6. Bibliografia

1. Case, D., Cheatham, T., Merz, K., Roitberg, A., Simmerling, C., Luo, R., Wang, J., Walker, R. & co-workers. (Maig 2018). Amber Home Page. Manuals. Recuperat de <http://ambermd.org/Manuals.php>.
2. Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Röthlisberger, D., Zanghellini, A., Gallaher, J. L., Tanaka, F., Barbas, C. F., Hilvert, D., Houk, K. N., Stoddard, B. L., & Baker, D. (2008). De novo computational design of retro-aldol enzymes. *Science*, 319(5868), 1387–1391.
3. Giger, L., Caner, S., Obexer, R., Kast, P., Baker, D., Ban, N., & Hilvert, D. (2013). Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nature Chemical Biology*, 9(8), 494–498.
4. Kiss, G., Çelebi-Ölçüm, N., Moretti, R., Baker, D., & Houk, K. N. (2013). Computational enzyme design. *Angewandte Chemie - International Edition*, 52(22), 5700–5725.
5. Obexer, R., Godina, A., Garrabou, X., Mittl, P. R. E., Baker, D., Griffiths, A. D., & Hilvert, D. (2017). Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nature Chemistry*, 9(1), 50–56
6. Roe, D. R. (Juliol 2014). AMBER: CPPTRAJ Tutorial C1. Introduction. Recuperat de <http://ambermd.org/tutorials/analysis/tutorial1/index.htm#Introduction>
7. Romero-Rivera, A., Garcia-Borràs, M., & Osuna, S. (2017). Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catalysis*, 7(12), 8524–8532.
8. Romero-Rivera, A., Maria-Solano, M. A. & Osuna, S. (2017). La importància de les eines computacionals en el disseny d'enzims d'interès industrial The importance of computational tools in the design of enzymes of industrial interest. *Revista de la Societat Catalana de Química*. 16, 29–37.
9. The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.