Université de Montréal

Modification de la spécificité de la transglutaminase par une approche semi-aléatoire

Un nouvel outil pour la synthèse des peptides

par

Roberto Antonio Chica

Département de chimie

Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (Ph. D.)
en chimie

Décembre, 2006

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée :

Modification de la spécificité de la transglutaminase par une approche semi-aléatoire

Un nouvel outil pour la synthèse des peptides

présentée par :

Roberto Antonio Chica

a été évaluée par un jury composé des personnes suivantes :

Dr Andreea R. Schmitzer, président-rapporteur

Dr Joelle N. Pelletier, directrice de recherche

Dr Jeffrey W. Keillor, codirecteur

Dr Robert Lortie, membre du jury

Dr Shahriar Mobashery, examinateur externe

Dr Andreea R. Schmitzer, représentant du doyen de la FES

Résumé

La synthèse chimique est la principale méthode employée pour produire les peptides en industrie. Plusieurs étapes d'activation et de protection/déprotection sont nécessaires, ce qui diminue l'efficacité globale de leur synthèse en plus de générer de nombreux déchets polluants. La synthèse enzymatique est une solution de rechange plus respectueuse de l'environnement. Elle réduit largement le besoin d'activation et de protection des groupements fonctionnels, en plus d'utiliser peu ou pas de solvants organiques polluants. Les enzymes utilisées actuellement pour la synthèse des peptides sont les protéases. Un des obstacles majeurs à l'utilisation de ces enzymes réside dans leur tendance à hydrolyser les peptides synthétisés, ce qui diminue le rendement. De plus, le problème de spécificité étroite des protéases empêche leur application à des réactions de couplage variées.

Pour s'attaquer à ces problèmes, nous proposons de développer de nouvelles enzymes catalysant de façon plus efficace et spécifique la synthèse des peptides. Ainsi, nous avons entrepris l'évolution dirigée, par une approche semi-aléatoire, de la transglutaminase (TGase), une enzyme qui catalyse déjà la formation (et non l'hydrolyse) des liens amides entre des peptides et des dérivés d'acides aminés. Nous voulons modifier sa spécificité dans le but de permettre la catalyse de la formation de liens amides entre une gamme d'acides α-aminés.

Pour ce faire, nous avons généré un modèle moléculaire de la TGase avec un substrat lié au site actif pour identifier les résidus impliqués dans la spécificité. Ces résidus ont été soumis à une mutagenèse semi-aléatoire pour générer cinq banques contenant chacune jusqu'à 8000 mutants. Une méthode d'expression de la TGase chez *Escherichia coli* a été développée de sorte à obtenir de l'enzyme soluble et active. Une méthode de criblage pour détecter les TGases mutantes ayant une spécificité désignée a été développée. Cette méthode, basée sur des analogues d'acides aminés dérivés avec un groupement fluorogène de type 7-hydroxycoumarine, a servi au criblage des cinq banques de mutants et

a permis d'identifier un mutant, Trp332Phe, ayant une différente spécificité par rapport à l'enzyme de type sauvage. Cette recherche a permis de développer un nouvel outil pour la synthèse des peptides qui a une efficacité catalytique similaire à celle de la papaïne, une protéase catalysant la synthèse des peptides.

Mots-clés : transglutaminase, enzyme, enzymologie, évolution dirigée, biocatalyse, modélisation moléculaire, cinétique, chimie bio-organique, chimie verte

Abstract

Chemical synthesis is the main method used to produce peptides in industry. It requires various activation, protection and deprotection steps which decrease overall yields and generate large quantities of waste that is harmful for the environment. Enzymatic peptide synthesis is an interesting alternative to chemical synthesis because it exploits the high selectivity of enzymes which require no protection and deprotection of potentially reactive functional groups, generating less waste. For this reason, enzymatic synthesis of peptides is an environmentally-friendly procedure. Proteases are currently the enzymes of choice for enzymatic peptide synthesis. However, they are prone to hydrolyse the peptide products, decreasing the yields. Furthermore, their narrow specificity prevents their application to the synthesis of various products.

The main goal of this work is to develop new catalysts for enzymatic peptide synthesis. These catalysts would be more efficient than proteases as they would not hydrolyse the peptide products formed. To reach our goal, we have undertaken to modify the specificity of a transglutaminase (TGase) by directed evolution in order to generate an enzyme that can recognize a variety of amino acids as donor or acceptor substrates. This would provide us with a novel tool for the synthesis of peptides as the bond formed by the enzyme would be peptidic rather than isopeptidic. The enzyme mechanism would not be altered by directed evolution since the catalyzed reaction would remain the same, namely the formation of an amide bond. The difference would be in the range of recognized substrates.

To modify the specificity of TGase by a semi-random approach, a protocol for the efficient expression and purification of the recombinant enzyme in *E. coli* was developped. Molecular modeling was applied to the generation of a model for the binding of peptide substrates on TGase, since no substrate-bound crystal structure of TGase is currently available. This model allowed identification of TGase residues that are involved in

substrate binding. These residues were subjected to combinatorial mutagenesis, thus generating five libraries of mutants, each containing up to 8000 mutants. Following screening of the mutants with a fluorimetric microplate-based assay that we developped, mutant Trp332Phe was identified as having a new desired specificity for the acyl-acceptor substrate. Mutant Trp332Phe can catalyze the formation of peptide bonds and represents a novel tool for the synthesis of peptides in aqueous solution.

Keywords : transglutaminase, TGase, enzyme, enzymology, directed evolution, biocatalysis, molecular modeling, kinetics, bioorganic chemistry, green chemistry

Table des matières

Liste des tableaux

CHAPITRE 5

CHAPITRE 6

Liste des figures

CHAPITRE 1

CHAPITRE 2

CHAPITRE 3

CHAPITRE 4

CHAPITRE 5

CHAPITRE 6

Liste des abréviations

7HC	7-hydroxycoumarine
Å	Ångström
Ac	groupement acétyle
ATP	adénosine triphosphate
bp	paire de base
BLAST	*Basic Local Alignment Search Tool*
Boc	*tert*-butyloxycarbonyle
BSA	albumine de sérum bovin
Cbz	carbobenzyloxycarbonyle
CVFF	*constant valence force field*
d	doublet
Da	Dalton
DCC	dicyclohexylcarbodiimide
dITP	désoxyinosine triphosphate
DMF	*N,N*-diméthylformamide
DMPDA	*N,N*-diméthyl-1,4-phénylènediamine
DNA	acide désoxyribonucléique
dpi	*dots per inch* (point par pouce)
DTT	dithiothréitol
E. coli	*Escherichia coli*
EC	*Enzyme Commission* (numéro de classification des enzymes)
EDTA	acide éthylènediaminetétraacétique
Enz	enzyme
epPCR	*error-prone polymerase chain reaction*
FAB	bombardement d'atomes rapides
FACS	*fluorescence-activated cell sorting*

FG	*functional group*
Fmoc	9-fluorénylméthylcarbamate
GABA	acide γ-aminobutyrique
GDH	glutamate déshydrogénase
GDP	guanosine diphosphate
GTP	guanosine triphosphate
GP	groupement protecteur
HBTU	2-(1H-benzotriazol-1-yl)-1,1,3,3-tétramethyluronium hexafluorophosphate
HOAt	1-hydroxy-7-azabenzotriazole
HOBt	1-hydroxybenzotriazole
HRMS	spectrométrie de masse à haute résolution
IMAC	*immobilized-metal affinity chromatography*
IPTG	isopropyl-β-D-thiogalactoside
LB	Luria-Bertani
m	multiplet
Mops	acide 3-(*N*-morpholino)propanesulfonique
M_r	masse moléculaire
NADH	nicotinamide adénine dinucléotide
NMR	résonance magnétique nucléaire
NTA	acide nitriloacétique
PCR	réaction en chaîne de la polymérase
PDB	*Protein data bank*
PNP	*p*-nitrophénolate
ppm	partie par million
ps	picoseconde
PyBOP	benzotriazolyloxy-tris[pyrrolidino]-phosphonium hexafluorophosphate

rmsd	*root mean square deviation*
rpm	tours par minute
s	singulet
SDS	dodécylsulfate de sodium
SDS-PAGE	électrophorèse sur gel de polyacrylamide avec dodécylsulfate de sodium
t	triplet
TGase	transglutaminase
Tris	tris(hydroxyméthyl)méthane
tRNA	acide ribonucléique de transfert
tTGase	TGase tissulaire
Xaa	acide aminé
λ_{em}	longueur d'onde d'émission
λ_{ex}	longueur d'onde d'excitation

A mi madre Alicia. No importa lo
que suceda, siempre estaremos juntos

Remerciements

Ces six dernières années passées au département de chimie de l'Université de Montréal m'ont permis de grandir énormément en tant que scientifique et en tant que personne. Durant ces années remplies d'efforts continus et de travail acharné, j'ai appris beaucoup de choses sur moi-même en plus de vivre plusieurs moments heureux et satisfaisants, et d'autres plus difficiles et douloureux. Néanmoins, ces années d'études doctorales en chimie ont été parmi les plus belles de ma vie. Tout ce que j'ai accompli durant ces six dernières années n'aurait pas été possible sans la contribution de nombreuses personnes proches de moi. J'aimerais leur manifester ma reconnaissance.

Je voudrais d'abord remercier mes parents, Alicia et Oscar. Sans l'énorme sacrifice qu'ils ont fait en quittant leur El Salvador natal à cause de la guerre civile pour venir recommencer leur vie à zéro au Canada, je ne serais pas ici en train de les remercier. J'aimerais rendre hommage à ma mère qui m'a enseigné la signification réelle des mots courage et espoir lors de sa lutte acharnée contre la leucémie... Elle a toujours été la plus grande admiratrice de mes succès scolaires, et j'ai bénéficié de son amour inconditionnel. J'aimerais saluer mon père parce qu'il m'a enseigné l'honnêteté et la ténacité, des valeurs importantes pour tout scientifique.

Je voudrais remercier ma compagne, Isabelle, pour son amour et son soutien, surtout lors de cette dernière année, où j'en ai eu particulièrement besoin. Un merci à mes frères Oscar et Javier et à ma sœur Patricia pour leur aide et leur amour.

J'aimerais témoigner ma très grande reconnaissance envers mes directeurs de recherche, Joelle et Jeff, aussi connus sous le nom de *Mom & Dad*. Ils ont été pour moi plus que des directeurs de recherche : ils ont été des amis. Merci beaucoup de m'avoir guidé dans mon apprentissage de la recherche scientifique et de m'avoir prêté votre assistance tout au long de mon doctorat, tout en me laissant beaucoup de liberté. Merci aussi de m'avoir montré que la recherche scientifique et la vie familiale n'étaient pas incompatibles.

Joelle, le temps fou que tu as consacré à m'aider à préparer des demandes de bourses, des présentations à des conférences et ma thèse a contribué énormément à mes réussites. Jeff, mon doctorat a été plus amusant grâce à notre complicité et à nos matchs de foot! Ce fut un énorme plaisir de travailler à vos côtés durant toutes ces années. Finalement, merci de m'avoir accordé le temps nécessaire pour que je m'occupe de mes responsabilités familiales.

Je m'en voudrais de passer sous silence les étudiants, postdocs, stagiaires et assistants de recherche des groupes Pelletier et Keillor avec qui j'ai eu la chance de travailler durant mon doctorat. Ils sont trop nombreux pour que je les nomme tous. Sachez que j'ai appris quelque chose de chacun d'entre vous et que vous avez contribué à créer une belle ambiance dans l'aile F-500. Un merci particulier à mes proches collaborateurs, Dr Steve « Sting » Gillet, Jessica Laroche et Farah-Jade Dryburgh avec qui j'ai eu le plaisir de travailler lors de mon projet de doctorat.

Je veux absolument souligner l'apport de mes amis et collègues Nicolas « Softy » Doucet, Jordan Volpato et Félix Doyon avec qui j'ai développé une grande camaraderie. Ensemble, nous avons découvert les études supérieures... en plus de faire de nombreuses folies! Je garderai toujours de bons souvenirs de nos *road trips*, de nos « routes des CD », de nos soirées de poker et du vocabulaire que nous avons développé et popularisé ensemble! C'est clé!

Un merci spécial à Marc Devleeschauwer, qui m'a initié à la recherche en chimie, et à Pierre-Yves De Wals pour avoir été comme un petit frère au labo.

J'aimerais aussi remercier le Conseil de recherche en sciences naturelles et en génie du Canada, le Fonds québécois de la recherche sur la nature et les technologies, Boehringer-Ingelheim ainsi que le département de chimie de l'Université de Montréal pour l'aide financière dont j'ai bénéficié.

CHAPITRE 1

Introduction

Depuis la synthèse de l'urée à partir du cyanate d'ammonium par Friedrich Wöhler en 1828 [1], la chimie organique a connu un essor continu et fulgurant. De nos jours, les chimistes sont capables de synthétiser une panoplie de molécules organiques ayant des propriétés diverses provenant de la variété quasi infinie des structures possibles pour ces composés à base de carbone et d'hydrogène. Que ce soit pour la synthèse de molécules telles des antibiotiques, des plastiques, des colorants, des additifs alimentaires, des agents de conservation et plusieurs autres composés d'utilité courante, les chimistes ont conçu plusieurs méthodes de synthèse organique utilisées couramment en recherche, mais aussi en industrie. La chimie organique contribue donc chaque jour à l'amélioration de notre qualité de vie.

Malgré son utilité évidente, la synthèse organique présente le désavantage de plus en plus problématique de générer des déchets nocifs pour l'environnement et pour la santé humaine. En effet, en 1993 aux États-Unis, l'industrie chimique a produit plus de 13 millions de tonnes de déchets, qui se sont retrouvés majoritairement dans l'air, l'eau et le sol [2], et ce, malgré les efforts de gestion efficace des déchets mis de l'avant par l'industrie. Pour s'attaquer à ce problème important de pollution environnementale par l'industrie chimique, les chimistes, à travers le monde, œuvrent de plus en plus à l'élaboration d'une chimie moins nocive pour l'environnement, qu'on appelle chimie verte.

1.1 La chimie verte

1.1.1 Définition

La chimie verte est une philosophie d'application et d'utilisation plutôt qu'une sous-discipline de la chimie. Elle se penche sur les problèmes de pollution liés intrinsèquement à l'application de la chimie par la conception de procédés et de produits chimiques permettant de réduire ou d'éliminer l'utilisation et la production de substances nocives pour l'être humain et pour l'environnement [3]. La chimie verte tente donc de réduire et de

prévenir la pollution à la source, c'est-à-dire dans les réactions et les procédés chimiques eux-mêmes. Elle touche plusieurs sous-disciplines de la chimie telles que la chimie organique, la chimie inorganique, la chimie physique, la chimie analytique, la biochimie et le génie chimique. Parmi les concepts principaux qu'elle préconise, on trouve la conception de procédés chimiques maximisant la quantité du matériau de départ se retrouvant dans le produit final, l'utilisation de solvants respectueux de l'environnement (en particulier l'eau), la conception de procédés énergétiquement efficaces et le développement de méthodes efficaces et préventives de gestion des déchets [2].

1.1.2 Les douze principes de la chimie verte

Pour aider à définir l'application de la philosophie de la chimie verte, Paul Anastas et John Warner, de l'Agence de la protection environnementale des États-Unis, ont élaboré douze principes qui résument ce qu'est la chimie verte et comment en faire l'application [4]. Les douze principes sont :

Limiter les déchets : Planifier les synthèses chimiques afin de prévenir la production de déchets qui devront être traités ou éliminés;

Concevoir des produits chimiques moins dangereux : Éviter de générer des produits chimiques toxiques;

Concevoir des synthèses chimiques moins dangereuses : Planifier des synthèses chimiques utilisant ou générant des produits peu ou pas nocifs pour l'environnement;

Utiliser des produits de départ renouvelables : Favoriser l'utilisation de produits de départ provenant de ressources renouvelables tels des produits dérivés de l'agriculture ou les déchets d'autres procédés chimiques plutôt que des produits dérivés du pétrole;

Privilégier les catalyseurs aux réactifs stoechiométriques : Les catalyseurs sont utilisés en très petite quantité et participent à la réaction à plusieurs reprises tandis que les réactifs stœchiométriques sont utilisés en excès et n'y participent qu'une fois;

Éviter autant que possible des dérivés chimiques tels les groupements protecteurs : Les groupements protecteurs requièrent des réactifs additionnels et génèrent plus de déchets;

Maximiser l'économie d'atomes : Concevoir les synthèses chimiques pour que les produits contiennent le plus possible d'atomes de départ, ce qui limite les pertes;

Utiliser des solvants à risques réduits : Éviter l'utilisation de solvants, d'agents de séparation et de tout autre produit chimique auxiliaire. Favoriser l'utilisation d'eau comme solvant;

Augmenter l'efficacité énergétique : Effectuer les réactions chimiques à température et pression ambiantes;

Concevoir des produits chimiques qui se dégradent après usage : Concevoir des produits qui se dégradent lorsqu'ils sont en contact avec des substances inoffensives, pour qu'ils ne s'accumulent pas dans la nature;

Faire les analyses en temps réel pour prévenir la pollution : Inclure un suivi en continu et en temps réel dans les synthèses chimiques pour minimiser et éliminer la formation de sous-produits de réaction;

Minimiser les risques d'accidents : Concevoir les produits chimiques et gérer leur état physique afin de minimiser les risques d'accidents tels les explosions, les incendies et les déversements dans l'environnement.

Ces douze principes forment un vaste éventail de recommandations touchant plusieurs aspects de la chimie moderne. Un aspect de la chimie moderne qui met en application plusieurs des principes énoncés ci-dessus est la biocatalyse.

1.1.3 La biocatalyse

La biocatalyse est l'utilisation d'enzymes et de microorganismes pour la synthèse de composés organiques utiles. Parmi les principes de la chimie verte qu'elle permet de mettre en application, on retrouve l'utilisation de catalyseurs biodégradables au lieu de réactifs

stœchiométriques et la substitution de solvants nocifs pour l'environnement par l'eau. De plus, puisque les enzymes et les microorganismes fonctionnent à des températures au-dessous de 100 °C et à pression atmosphérique, la biocatalyse permet d'augmenter l'efficacité énergétique des réactions catalysées.

L'utilité de la biocatalyse dans une application de la chimie verte réside dans la très haute sélectivité dont font preuve les biocatalyseurs comme les enzymes. Ces biomolécules démontrent trois types de sélectivité : la chimiosélectivité, la régiosélectivité et l'énantiosélectivité [5]. La chimiosélectivité des enzymes leur permet d'agir sur une seule fonction chimique d'une molécule contenant plusieurs fonctions différentes qui réagiraient néanmoins avec le même réactif chimique. La régiosélectivité des enzymes est leur capacité de réagir spécifiquement sur une unique fonction chimique en présence d'autres fonctions chimiques identiques sur la même molécule. L'énantiosélectivité des enzymes découle de leur chiralité intrinsèque puisque ces biocatalyseurs sont formés exclusivement d'acides aminés ayant une configuration L. Les enzymes peuvent donc différencier les énantiomères entre eux et réagir seulement avec un seul. Ces trois types de sélectivité des enzymes, en particulier l'énantiosélectivité, en font des catalyseurs très utiles pour la synthèse organique. De plus, la haute sélectivité des enzymes présente l'avantage de générer très peu de sous-produits dont il faut se débarrasser par la suite, en plus de nécessiter peu ou pas de groupements protecteurs. Donc, la haute sélectivité et la biodégradabilité des biocatalyseurs en font des outils de choix dans l'application de la chimie verte.

Malgré l'utilité évidente des enzymes dans la biocatalyse, plusieurs problèmes courants doivent encore être résolus pour qu'on bénéficie pleinement des avantages procurés par ces biocatalyseurs. Par exemple, les enzymes ne sont pas toujours stables ou assez actives pour être de bons biocatalyseurs sous les conditions des procédés industriels. De plus, le biocatalyseur approprié pour la réaction désirée peut ne pas exister dans la nature ou ne pas avoir été déjà isolé. Néanmoins, depuis quelques décennies, les biocatalyseurs ont été utilisés avec succès dans la synthèse de produits chimiques courants

et spécialisés dans les industries chimique, pharmaceutique et agroalimentaire [6, 7]. Par exemple, les biocatalyseurs sont utilisés dans la synthèse industrielle à grande échelle de l'aspartame, de l'acrylamide, de l'éphédrine, de l'acide aspartique et de pénicillines [3]. De plus, ils remplacent déjà les catalyseurs chimiques traditionnels dans les processus de blanchiment des pâtes et papiers [3] et d'hydrogénation des gras insaturés lors de la préparation de la margarine [8]. Ils présentent donc une utilité économique en plus de contribuer à l'application de la chimie verte. Une application supplémentaire de la biocatalyse est la synthèse des peptides.

1.2 La synthèse des peptides

Les peptides sont des biomolécules formées d'acide aminés qui, dans la nature, jouent plusieurs rôles physiologiques importants [9]. Ceci fait en sorte que les peptides ont des applications variées et très en demande, en particulier dans les industries pharmaceutique et agroalimentaire [10]. Par exemple, de nombreux composés pharmaceutiques, tels des inhibiteurs d'enzymes, des hormones, des antibiotiques, des antiviraux et des neurotransmetteurs, sont des peptides ou sont basés sur des peptides biologiquement actifs [11-13]. De plus, certains dérivés peptidiques non naturels, comme l'aspartame et autres succédanés de sucre hypocaloriques, figurent parmi les produits de synthèse industrielle de premier plan [10, 14]. Pour ces raisons, l'importance de la synthèse des peptides est en pleine croissance en recherche fondamentale tout comme dans l'industrie.

1.2.1 La synthèse chimique

La synthèse chimique des peptides est la principale méthode de production de peptides à l'échelle industrielle. Pour synthétiser un peptide, il faut procéder au couplage du groupement carboxyle (C-terminus) d'un acide aminé et du groupement amino (N-terminus) d'un deuxième acide aminé. Cette réaction de couplage nécessite l'activation

du groupement carboxyle et se fait avec des agents de couplage tels les carbodiimides et les oximes aromatiques. Les carbodiimides, comme le dicyclohexylcarbodiimide (DCC), ont été les premiers agents de couplage peptidique développés. Ils activent les fonctions carboxyles des acides aminés en permettant la formation d'urées *O*-acylées hautement réactives (Figure 1A). Le désavantage des carbodiimides est qu'ils sont trop réactifs ce qui cause la racémisation des acides aminés [15]. Pour pallier ce problème, les oximes aromatiques tels le 1-hydroxybenzotriazole (HOBt) et le 1-hydroxy-7-azabenzotriazole (HOAt) ont été développés. Ces agents de couplage réagissent avec les urées *O*-acylées synthétisées par les carbodiimides et forment des esters activés moins réactifs (Figure 1B), ce qui diminue la possibilité de racémisation [15]. De nouvelles méthodes d'activation ne requérant plus l'utilisation des carbodiimides ont aussi été développées. L'ester activé est introduit en tant que sel d'uronium ou de phosphonium d'un anion non nucléophile tel le tétrafluoroborate ou l'hexafluorophosphate. Des agents de couplage de ce genre sont le HBTU (2-(1H-benzotriazol-1-yl)-1,1,3,3-tétramethyluronium hexafluorophosphate) [16] et le PyBOP (benzotriazolyloxy-tris[pyrrolidino]-phosphonium hexafluorophosphate) [17].

En plus de l'activation de la fonction carboxyle lors de la synthèse des peptides, il faut procéder à la protection chimique de divers groupes fonctionnels que l'on veut conserver intacts (-OH, -SH, -COOH, $-NH_2$, etc.) et qui se retrouvent sur les chaînes latérales des acides aminés. Plusieurs groupements protecteurs ont été développés à cette fin, comme le groupement Fmoc (9-fluorénylméthylcarbamate), le groupement Boc (*tert*-butoxycarbonyle) et le groupement Cbz (carbobenzyloxy). Évidemment, l'utilisation de groupements protecteurs implique des étapes supplémentaires lors de la synthèse des peptides, ce qui génère plus de déchets et entraîne l'achat de solvants et de réactifs supplémentaires, engendrant des coûts additionnels. De plus, il faut une méthode d'élimination et de gestion sécuritaire des solvants usés.

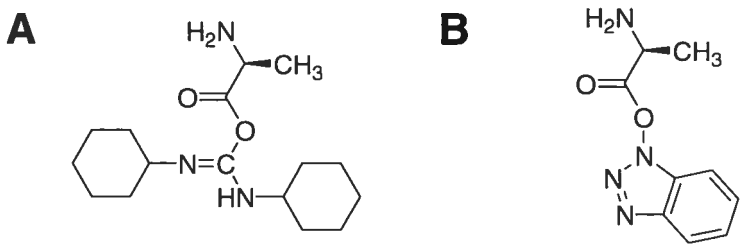


Figure 1. Activation de la fonction carboxyle des acides aminés lors de la synthèse peptidique. (A) Urée *O*-acylée formée par la réaction du DCC avec l'alanine. (B) Ester activé formé par la réaction de l'urée *O*-acylée **A** avec le HOBt.

À la suite du couplage chimique, le cycle de protection, d'activation, de réaction et de déprotection est répété pour allonger le peptide. Les rendements imparfaits obtenus à chaque étape donnent vite lieu à une gamme de sous-produits indésirables. Notamment, des sous-réactions de l'acide aminé activé avec l'eau, avec les groupes fonctionnels des chaînes latérales d'autres acides aminés ou encore avec des bases diminuent l'efficacité et/ou la pureté énantiomérique du produit final. En dernier lieu, il existe le problème d'isolation et de purification du produit après chaque étape afin d'en éliminer les excès de réactifs, les groupements protecteurs clivés et les autres sous-produits. Ce problème est partiellement résolu par la synthèse sur support solide, qui facilite les étapes d'isolation et de purification.

La synthèse sur support solide est la méthode de choix pour la synthèse chimique des peptides. Toutefois, cette méthode génère beaucoup de déchets, car, pour qu'elle donne de bons rendements, il faut utiliser un excès de substrats et il faut laver la résine avec beaucoup de solvant, sans compter qu'il faut quand même utiliser des groupements protecteurs et des agents de couplage. La synthèse chimique des peptides est donc un procédé polluant qui ne respecte pas les douze principes de la chimie verte. Des alternatives à la synthèse chimique des peptides plus respectueuses de l'environnement sont la biosynthèse et la synthèse enzymatique.

1.2.2 La biosynthèse

La biosynthèse est la synthèse de produits chimiques ayant lieu à l'intérieur d'organismes vivants tels les microorganismes et les animaux transgéniques. Quoique utile, la biosynthèse des peptides ne permet pas la synthèse de peptides formés d'acides aminés non naturels. De plus, la production microbienne de petits peptides est limitée par la présence de peptidases qui détruisent une partie des peptides produits, ce qui diminue les rendements. Il peut aussi être difficile d'obtenir un produit de pureté satisfaisante à un coût raisonnable avec cette méthode. Enfin, l'utilisation d'organismes transgéniques peut présenter des problèmes éthiques, en plus de compliquer l'application à grande échelle de cette méthode. Une alternative plus prometteuse est la synthèse *in vitro* à l'aide d'enzymes.

1.2.3 La synthèse enzymatique

La synthèse enzymatique des peptides réduit largement les besoins d'activation et de protection chimique des acides aminés et assure la pureté énantiomérique des produits en exploitant la très haute sélectivité des enzymes. Les deux principales approches de cette méthode sont l'utilisation de peptidyltransférases, et la mise en œuvre de protéases. La première approche implique l'utilisation *in vitro* de composantes isolées de la machinerie cellulaire [18]: ribosomes, ARN messager et de nombreux ARN de transfert et autres facteurs cellulaires, ainsi que des substrats coûteux comme l'ATP. Cette méthode est très dispendieuse et demeure difficile d'application à grande échelle.

L'utilisation de protéases pour la synthèse des peptides est la méthode biocatalytique la plus répandue, étant déjà utilisée pour la synthèse à l'échelle industrielle de l'aspartame et de l'insuline humaine [19]. La synthèse enzymatique de peptides à l'aide de protéases peut se faire par contrôle thermodynamique ou cinétique (Figure 2). Dans une réaction à contrôle thermodynamique, on exploite l'inversion de la capacité des protéases à hydrolyser les liaisons peptidiques. Par le contrôle des conditions réactionnelles, on peut

modifier l'équilibre et favoriser la condensation des liaisons peptidiques au lieu de l'hydrolyse. Plusieurs stratégies ont été proposées pour ce faire [14, 18]. La plus commune est de favoriser la précipitation du produit de condensation. Par exemple, en utilisant des réactifs de départ générant un produit de condensation peu soluble dans le milieu réactionnel aqueux, ce produit précipitera lors de sa synthèse, ce qui déplacera l'équilibre vers la réaction de condensation. Si le produit de condensation est soluble dans l'eau, on peut aussi utiliser des cosolvants non aqueux pour favoriser sa précipitation. L'utilisation de systèmes biphasiques et de micelles permet aussi de déplacer l'équilibre vers la réaction de condensation en séquestrant le produit de condensation dans une phase immiscible dans l'eau ou dans une micelle.

Dans une réaction à contrôle cinétique, on joue sur les rapports de compétition entre l'hydrolyse et la condensation. En utilisant un substrat activé, les protéases peuvent agir en tant que transférases et catalyser le transfert d'un groupement acyle sur un nucléophile [14, 19]. En présence d'une amine, ce transfert d'acyle se fera sur ce nucléophile plutôt que sur l'eau, favorisant ainsi la réaction de condensation aux dépends de l'hydrolyse. Ce type de réaction n'est possible qu'avec des protéases à sérine et à cystéine. Une des limitations majeures de l'utilisation de protéases pour la synthèse enzymatique des peptides réside dans la difficulté d'éliminer complètement l'hydrolyse des liens peptidiques déjà formés, ce qui nuit au rendement global lors de la production de peptides d'une certaine taille. De plus, le problème de spécificité étroite propre aux protéases subsiste. Pour aider à résoudre ces problèmes, une classe d'enzymes, les transglutaminases, pourrait être envisagée comme catalyseur de la synthèse peptidique.

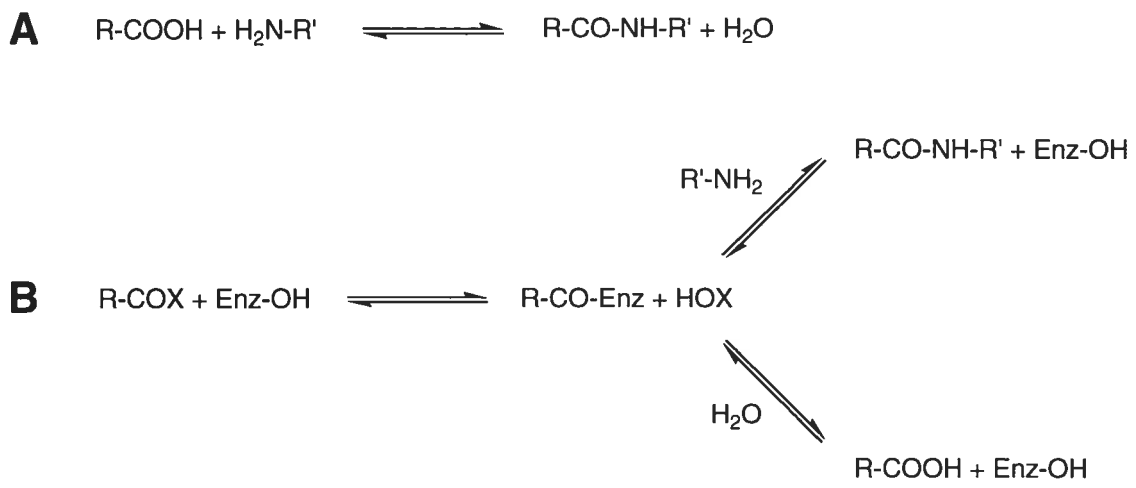


Figure 2. Synthèse enzymatique des peptides catalysée par les protéases. (A) Synthèse sous contrôle thermodynamique. (B) Synthèse sous contrôle cinétique [18].

1.3 Les transglutaminases

Les transglutaminases (TGases) composent une famille d'enzymes faisant partie de la classe des transférases (EC 2.3.2.13), qui catalysent la réticulation post-traductionnelle des protéines. La réticulation des protéines par les TGases génère des produits de haute masse moléculaire très résistants à la dégradation protéolytique et au stress mécanique. Ces produits se retrouvent dans plusieurs tissus et forment en partie la peau, les cheveux, les matrices extracellulaires et les caillots sanguins [20]. Les TGases sont impliquées dans plusieurs rôles physiologiques importants telles l'endocytose, la coagulation sanguine, la formation de l'épiderme, l'apoptose et la régulation de la croissance cellulaire [21-23]. Une mauvaise régulation de l'activité des TGases mène à plusieurs désordres physiologiques tels la formation des cataractes [24], la maladie cœliaque [25] et le psoriasis [26], et semblerait être impliquée dans le développement de maladies neurodégénératives telles les maladies d'Alzheimer et de Huntington [27].

1.3.1 Réaction catalysée et mécanisme catalytique

Les TGases réticulent les protéines en catalysant une réaction de transfert d'acyle dépendante du Ca^{2+} entre le groupement γ-carboxamide d'un résidu Gln et le groupement ε-amino d'un résidu Lys, ce qui mène à la formation d'un lien isopeptidique γ-glutamyl-ε-lysine (Figure 3). Ces liaisons isopeptidiques ne sont pas susceptibles de se faire protéolyser. En plus de catalyser la formation de liens isopeptidiques entre des peptides et des protéines, les TGases peuvent aussi catalyser d'autres modifications post-traductionnelles telles l'incorporation d'amines dans les protéines et la désamination sur des sites spécifiques de résidus Gln (Figure 4).

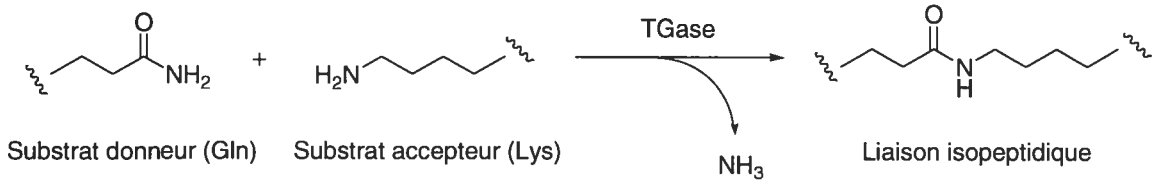


Figure 3. Réaction catalysée par les TGases.

Les réactions catalysées par les TGases procèdent *via* un mécanisme de type « ping-pong modifié » [28] (Figure 5). La réaction est effectuée par une triade catalytique Cys-His-Asp conservée, similaire à celle des protéases à cystéine. Il est généralement admis que les résidus Cys et His forment une paire d'ions thiolate-imidazolium [29, 30]. La première étape du mécanisme est l'étape d'acylation, où une protéine ou un peptide contenant un résidu Gln exposé, agissant comme substrat donneur d'acyle, réagit avec la Cys catalytique, formant un intermédiaire acyl-enzyme covalent et libérant de l'ammoniaque. Par la suite, cet intermédiaire acyl-enzyme réagit avec le second substrat, l'accepteur d'acyle, et lui transfère le groupement γ-glutamyle, régénérant par le fait même l'enzyme libre : c'est l'étape de désacylation. Lorsque le substrat accepteur est une amine, l'étape de désacylation est appelée transamidation. Pour la transamidation, le substrat accepteur peut être presque

n'importe quelle amine primaire mais *in vivo*, il s'agit habituellement d'un résidu Lys d'une protéine ou d'un peptide. En l'absence d'amine primaire, l'intermédiaire acyl-enzyme peut aussi être hydrolysé, transformant le résidu Gln du substrat donneur en résidu Glu, mais à une vitesse inférieure par rapport à la transamidation. L'étape de désacylation se nomme alors hydrolyse.

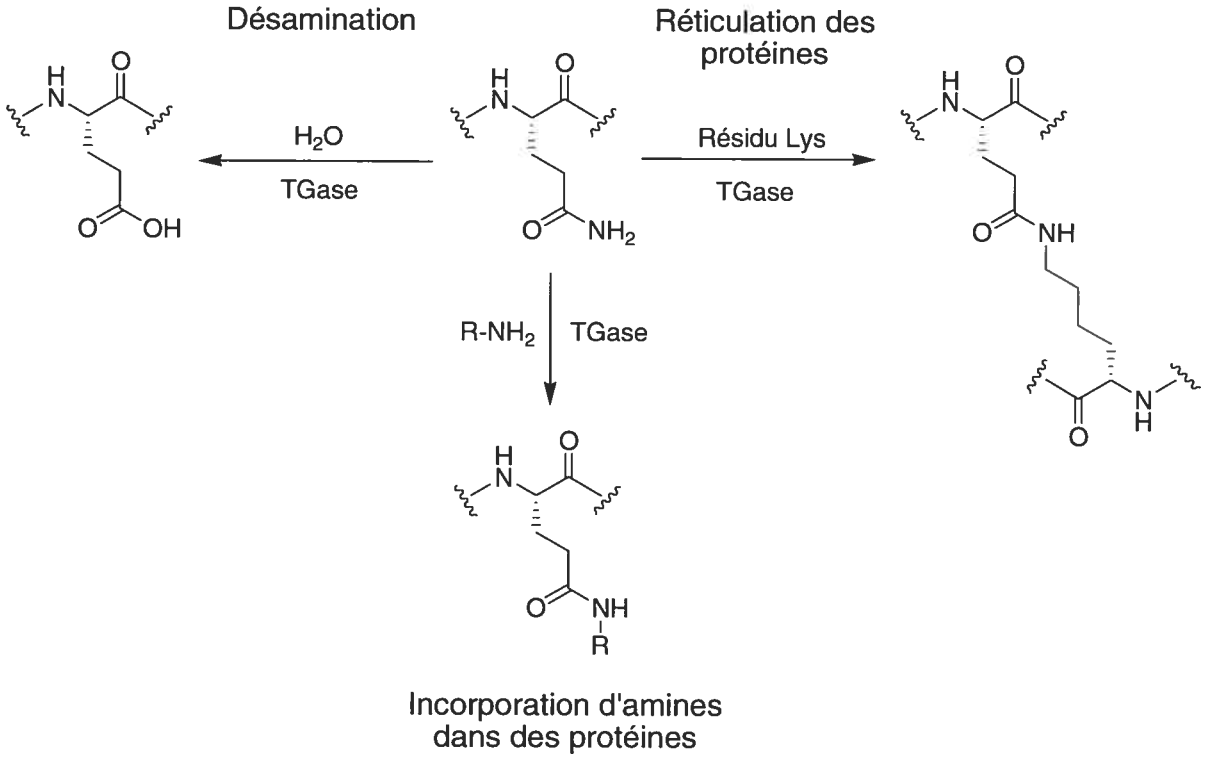


Figure 4. Modifications post-traductionnelles des protéines catalysées par les TGases.

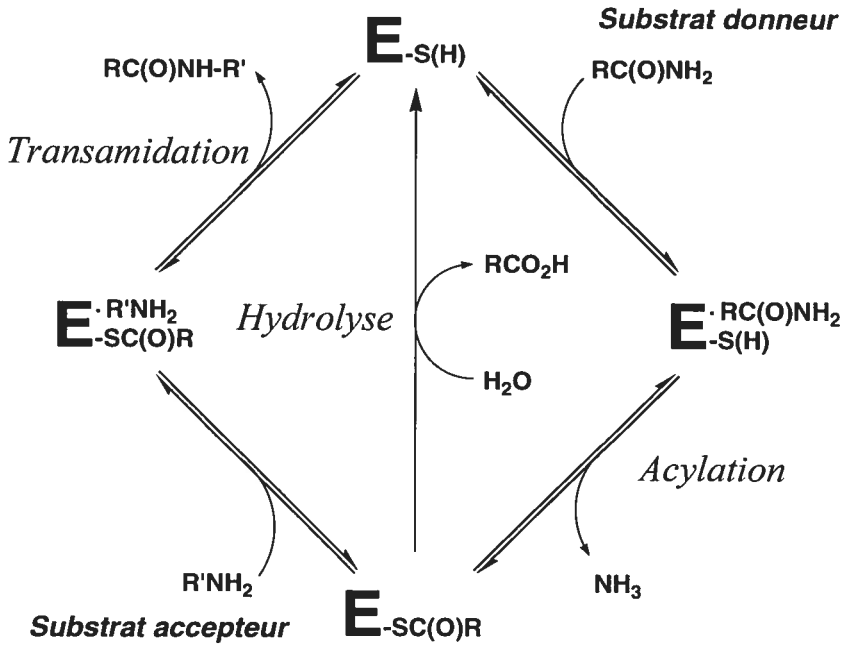


Figure 5. Mécanisme de type « ping-pong modifié » des TGases.

L'étape limitante du mécanisme de la TGase est l'étape de désacylation (transamidation ou hydrolyse) [31, 32]. En effet, il a été démontré que, même en utilisant des amines hautement nucléophiles comme substrats accepteurs, la vitesse de la désacylation restait inférieure à la vitesse d'acylation [32]. Ceci démontre que les TGases sont très efficaces pour activer le transfert du groupement γ-glutamyl vers les amines primaires, aux dépens de l'eau, et ce, bien que l'eau soit à une concentration beaucoup plus élevée dans le milieu réactionnel.

1.3.2 Classification

Les TGases sont des transférases de classe EC 2.3.2.13 (protéine-glutamine γ-glutamyltransférases) qui sont présentes chez tous les vertébrés de même que chez les bactéries [33], les plantes [34], les levures [35] et les nématodes [36]. Chez les mammifères, neuf TGases ont été identifiées par séquençage génomique (Tableau 1) [20,

23, 37-39] et ont été classifiées selon leur séquence. Cependant, seulement six TGases ont été caractérisées au niveau protéique : il s'agit du facteur XIIIa, de la TGase de kératinocyte, de la TGase tissulaire, de la TGase épidermale, de la TGase de prostate et de la TGase X.

Tableau 1. Les différentes TGases de mammifères

Type de TGase	Synonyme	Masse approximative (kDa)	Localisation cellulaire	Expression tissulaire
Facteur XIIIa	TGase du plasma	83 (monomère)	Extracellulaire et cytosolique	Sang
TGase 1	TGase de kératinocyte	90	Majoritairement membranaire	Épithélium
TGase 2	TGase tissulaire	80	Cytosolique, nucléaire et extracellulaire	Ubiquitaire
TGase 3	TGase épidermale	77	Cytosolique	Épithélium
TGase 4	TGase de prostate	77	Extracellulaire	Prostate
TGase 5	TGase X	81	Inconnue	Ubiquitaire
TGase 6	TGase Y	Inconnue	Inconnue	Inconnue
TGase 7	TGase Z	80	Inconnue	Ubiquitaire
Bande 4.2	B4.2 (inactive)	77	Membranaire	Sang

Le facteur XIII, aussi connu comme TGase du plasma, est le dernier zymogène activé lors de la cascade de coagulation sanguine et participe donc à la formation du caillot sanguin. Cette protéine tétramérique est composée de deux chaînes A formant la partie transglutaminase et de deux chaînes B n'ayant aucune activité catalytique [40]. Elle est activée à la suite de la libération du peptide d'activation de la chaîne A par l'action de la

thrombine, générant ainsi le facteur XIIIa. Le facteur XIIIa est la seule TGase de mammifère qui est homodimérique : chaque monomère est formé de 732 résidus et a une masse moléculaire de 83 kDa. Cette TGase peut se retrouver dans le plasma et dans le cytoplasme de plusieurs types de cellules.

La TGase de kératinocyte, ou TGase de type 1, est une enzyme monomérique de 90 kDa qui se retrouve liée, *via* une ancre lipidique, au côté cytoplasmique de la membrane cellulaire des kératinocytes. C'est la plus grande TGase (817 résidus, 90 kDa), car elle possède une région d'ancrage à la membrane en N-terminal, qui contient un groupe de cinq cystéines sur lesquelles des acides gras tels l'acide myristique et l'acide palmitique peuvent former des liens thioester, donnant lieu à l'ancre lipidique [41]. Ses rôles biologiques sont la différenciation des kératinocytes et la formation de la couche cornée de la peau. La TGase de kératinocyte est protéolysée à deux endroits, ce qui l'active, et les trois fragments ainsi générés restent associés entre eux [42].

La TGase tissulaire (TGase de type 2), une protéine monomérique de 690 résidus (80 kDa), est le membre le plus répandu et le plus étudié de cette famille. Cette TGase se retrouvant dans tous les tissus est majoritairement cytosolique, mais peut également se retrouver dans le noyau [43] et dans le milieu extracellulaire [37]. Elle possède une activité d'hydrolyse de nucléotides (GTP et ATP), ce qui en fait un membre de la famille des protéines G [44]. La liaison au GTP inhibe son activité transférase. Cette TGase démontre de plus des activités disulphide isomérase [45] et kinase [46], et contient des sites de liaison à la fibronectine et à l'intégrine [47]. Le rôle physiologique précis de cette TGase n'est pas encore clairement démontré, mais elle semble être impliquée dans l'induction de l'apoptose [48], l'endocytose, l'adhésion cellulaire, le développement de la matrice extracellulaire et la différenciation cellulaire [23, 37].

La TGase épidermale (TGase de type 3) est une TGase monomérique impliquée dans la différenciation des kératinocytes et dans la formation de l'enveloppe cellulaire dans l'épiderme et le follicule pileux. Cette enzyme consiste en deux chaînes polypeptidiques

qui sont exprimées en une seule proenzyme de 692 résidus (77 kDa) requérant une protéolyse pour être activée. À la suite de cette protéolyse, deux fragments globulaires qui demeurent associés sont générés [37]. La TGase épidermale, comme la TGase tissulaire, peut aussi lier et hydrolyser le GTP, mais le rôle de cette activité demeure pour l'instant inconnu [49].

Les autres TGases de mammifères sont encore peu étudiées. La TGase de prostate (TGase de type 4) est impliquée dans la formation du caillot post-coïtal chez le rat. Son rôle chez l'humain est inconnu. La TGase X (TGase de type 5) contribue à la formation de la couche cornée des kératinocytes et peut lier le GTP [50]. La protéine Bande 4.2 de l'érythrocyte fait partie de la famille des TGases mais n'a pas d'activité enzymatique, car la Cys catalytique est remplacée par une Ala. C'est un composant majeur du cytosquelette des érythrocytes et elle joue un rôle important dans la structure et les propriétés mécaniques de ces cellules [37]. Le rôle physiologique des TGases de type 6 et 7 n'a pas encore été élucidé.

Une TGase microbienne provenant de *Streptoverticillium mobaraense* a aussi été étudiée. C'est la seule TGase ne provenant pas de vertébrés à avoir été isolée. Cette TGase est très différente des TGases de mammifères, ne possédant aucune homologie de séquence ni de structure avec ses homologues de mammifères. C'est une enzyme sécrétée, de 331 résidus (38 kDa), qui nécessite le clivage d'un propeptide pour être activée. Cette enzyme possède une triade catalytique Cys-His-Asp et catalyse la même réaction que les autres TGases mais ne requiert pas de Ca^{2+} pour être active, ce qui la distingue des TGases provenant d'eucaryotes. Son rôle biologique n'est pas connu [51].

1.3.3 Structure

Les structures cristallines de quatre TGases de vertébrés appartenant à trois différentes classes ont été résolues. Il s'agit de la TGase de plasma humaine (facteur XIIIa) [30, 52-54], des TGase tissulaires humaine [55] et de poisson (*Pagrus major*) [56], de

même que de la TGase épidermale humaine [49, 57-59]. Malgré des différences significatives dans leur structure primaire, ces TGases adoptent toutes une structure tridimensionnelle similaire. Ces structures sont formées de quatre domaines (Figure 6A) : le domaine N-terminal sandwich β, le noyau catalytique α/β, le tonneau β 1 et le tonneau β 2 C-terminal. Le domaine N-terminal sandwich β a été démontré comme étant le site de liaison de la fibronectine et de l'intégrine chez la TGase tissulaire. Le noyau catalytique α/β est composé d'hélices α et de feuillets β et contient le site actif pour l'activité transférase, mais aussi le site actif pour l'activité d'hydrolyse de nucléotides des TGases de types 2, 3 et 5. Ce domaine contient aussi le site de liaison au calcium, dont la localisation demeure floue pour la TGase tissulaire. D'après des études de délétions, les domaines tonneaux β 1 et 2 ne semblent pas être indispensables pour l'activité catalytique de la TGase tissulaire [60] et du facteur XIIIa [61]. Les TGases de vertébrés contiennent plusieurs résidus Cys, mais aucun pont disulphure n'a été identifié [62].

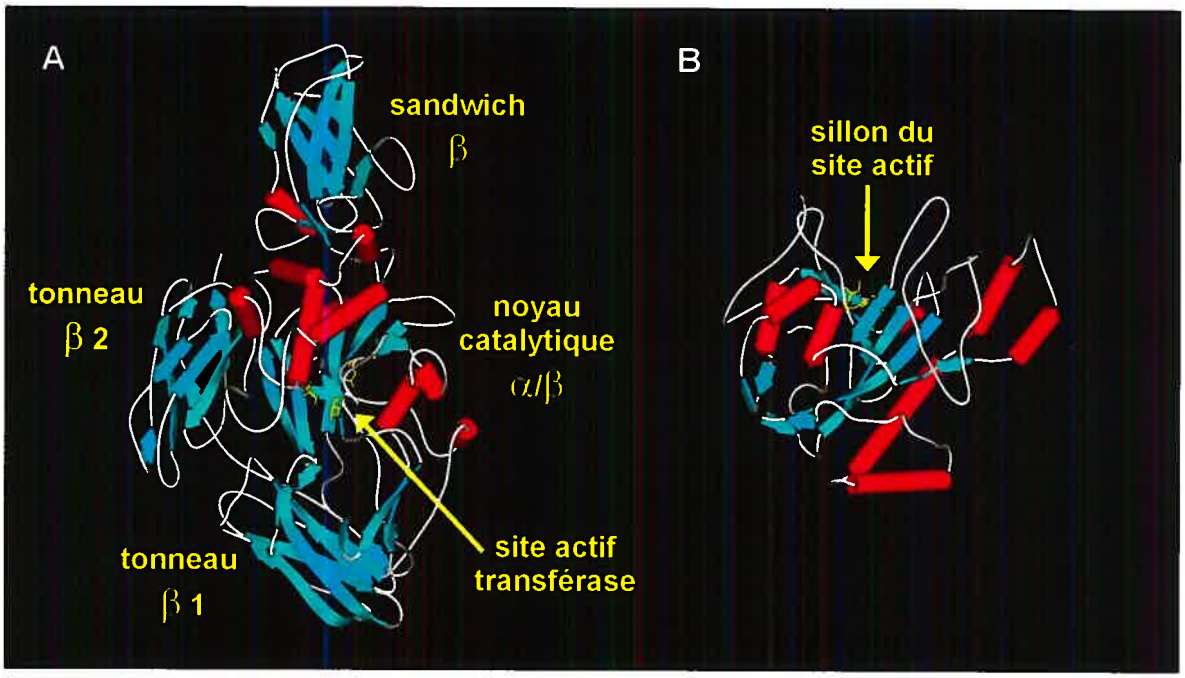


Figure 6. Structure cristalline de TGases. (A) TGase tissulaire de *Pagrus major* (Code PDB 1G0D). (B) TGase microbienne de *Streptoverticillium mobaraense* (Code PDB 1IU4).

La structure cristalline de la TGase microbienne provenant de *Streptoverticillium mobaraense* a aussi été résolue [51]. La structure de cette enzyme est complètement différente de celle des TGases de mammifères et ne présente pas d'homologie de structure avec aucune autre protéine connue. Cette enzyme est formée d'un seul domaine en forme de coeur contenant des hélices α et des feuillets β (Figure 6B). Le site actif se retrouve à la base d'un sillon de ~16 Å de profondeur contenant le seul résidu Cys de l'enzyme.

1.3.4 Spécificité

Les substrats naturels des TGases sont des protéines et des peptides. Ces substrats des TGases peuvent être divisés en deux catégories : les substrats donneurs d'acyle possédant une Gln réactive et les substrats accepteurs d'acyle possédant une Lys réactive. À ce jour, plusieurs protéines ont été identifiées comme étant des substrats de TGases [63]. Puisque les protéines peuvent contenir à la fois des résidus Gln et Lys, certains des substrats des TGases peuvent être à la fois un substrat donneur et un substrat accepteur. Il ne semble pas y avoir de séquence consensus reconnue par les TGases, car celles-ci peuvent réagir sur presque n'importe quel résidu Gln ou Lys accessible. Par contre, Khosla et ses collègues [64, 65] ont démontré que la séquence PQPQLPY contenue dans les gliadines est une séquence de haute affinité chez les substrats donneurs de la TGase tissulaire pour la réaction de désamination. Les TGases peuvent aussi reconnaître plusieurs mono- et polyamines naturelles avec une spécificité large. En effet, plusieurs amines telles la putréscine, la spermidine, la spermine, la cadavérine, l'histamine et la sérotonine sont utilisées par les TGases comme accepteurs d'acyle *in vitro* [63, 66].

La spécificité des TGases envers des substrats non naturels a surtout été étudiée chez la TGase tissulaire (Tableau 2). Ces substrats non naturels incluent des peptides synthétiques et plusieurs amines primaires. La TGase tissulaire démontre une très large spécificité envers le substrat accepteur d'acyle. Ce substrat peut être presque n'importe quelle amine primaire [67]. Des amines primaires aliphatiques, telles la glycinamide, la

glycine méthyl ester et la cadavérine [32, 66, 68], de même que des anilines, telle la *N,N*-diméthyl-1,4-phénylènediamine [69], peuvent être des substrats de l'enzyme. Toutefois, les amines contenant des fonctions acides carboxyliques libres, tels les acides aminés, et les amines ayant des groupements volumineux adjacents au groupement amino, tel l'amide aminé L-tyrosinamide, ne sont pas des substrats [66].

Tableau 2. Spécificité de la TGase tissulaire pour les substrats non naturels

Substrats donneurs		Substrats accepteurs	
Molécules reconnues	Molécules non reconnues	Molécules reconnues	Molécules non reconnues
Peptide contenant L-Gln	Peptide contenant D-Gln ou L-Asn	Amines primaires	Amines secondaires
	L-Gln	Anilines	Acides aminés
Peptides contenant un ester aromatique de L-Glu	Peptide contenant un dérivé anilide de L-Gln	Dérivés d'acides aminés sans fonction carboxylate libre et ayant de petites chaînes latérales	Dérivés d'acides aminés sans fonction carboxylate libre et ayant des chaînes latérales volumineuses
	Amides secondaires	Eau	

La spécificité des TGases tissulaires pour les substrats donneurs est plus étroite (Tableau 2). Les peptides contenant de la L-Gln sont reconnus, ce qui n'est pas le cas de peptides contenant les résidus similaires L-Asn et D-Gln. De plus, pour qu'un peptide soit un substrat de la TGase, il faut qu'il ait minimalement deux résidus et le groupement protecteur Cbz en N-terminal. La L-Gln seule n'est pas un substrat des TGases [68]. Pour

cette raison, le peptide synthétique *N*-carbobenzyloxyglutaminylglycine (Cbz-Gln-Gly) est le substrat donneur synthétique le plus couramment utilisé pour les études cinétiques des TGases tissulaires. Ce composé a un K_M élevé, d'environ 3 mM [70-72]. Il est aussi connu que les esters aromatiques γ-glutamyles sont de bons substrats des TGases. Par exemple, le substrat artificiel *N*-carbobenzyloxy-L-glutamyl(γ-*p*-nitrophényl ester)glycine [32] est utilisé en analyses cinétiques et a un K_M de 20 μM, ce qui est beaucoup plus bas que son analogue Cbz-Gln-Gly. Par contre, les anilides et les amides secondaires ne sont pas des substrats [73], ce qui démontre que la fonction γ-carboxamide de la L-Gln est le seul amide reconnu par l'enzyme. La TGase tissulaire peut donc reconnaître plusieurs dipeptides, tels Cbz-Gln-Gly, comme substrats donneurs et plusieurs dérivés d'acides aminés, telle la glycinamide, comme substrats accepteurs.

1.4 Description du projet de recherche

1.4.1 Objectif

L'objectif général du projet est de développer de nouvelles enzymes catalysant de façon plus efficace et économiquement avantageuse la synthèse des peptides afin de contribuer au développement de la chimie verte au Canada. L'objectif spécifique est de modifier la spécificité de la TGase pour la rendre capable de catalyser le couplage entre la fonction amide d'un amide aminé protégé (GP-Xaa-NH_2) comme substrat donneur d'acyle et le groupe α-amino de la forme amide ou ester d'un second acide aminé (Xaa-NH_2 ou Xaa-OMe) comme substrat accepteur d'acyle (Figure 7).

L'avantage d'utiliser les TGases plutôt que les protéases pour la synthèse enzymatique des peptides est qu'elles génèrent déjà des liaisons de type amide entre des peptides et des dérivés d'acides aminés en plus de n'avoir aucune action hydrolytique envers les produits de la réaction, ce qui pourrait donner de meilleurs rendements. De plus, en utilisant la TGase comme enzyme de départ, la modification proposée n'impliquera

aucun changement au mécanisme catalytique de l'enzyme : l'enzyme de type sauvage catalyse la formation des liens isopeptidiques à partir d'amides et d'amines primaires tandis que l'enzyme modifiée catalysera la formation des liens peptidiques à partir de ces mêmes classes de composés. La spécificité de la TGase pour les substrats accepteurs étant déjà large, il devrait être possible de l'élargir encore de sorte à ce qu'elle utilise une plus grande gamme de dérivés d'acides aminés comme substrats. De même, il est aussi envisageable de modifier la spécificité de la TGase pour le substrat donneur de sorte à ce que cette enzyme reconnaisse une fonction α-carboxamide au lieu de γ-carboxamide. Ce faisant, nous pourrons obtenir une enzyme catalysant la synthèse des liaisons peptidiques à partir d'une grande variété de dérivés d'acides aminés.

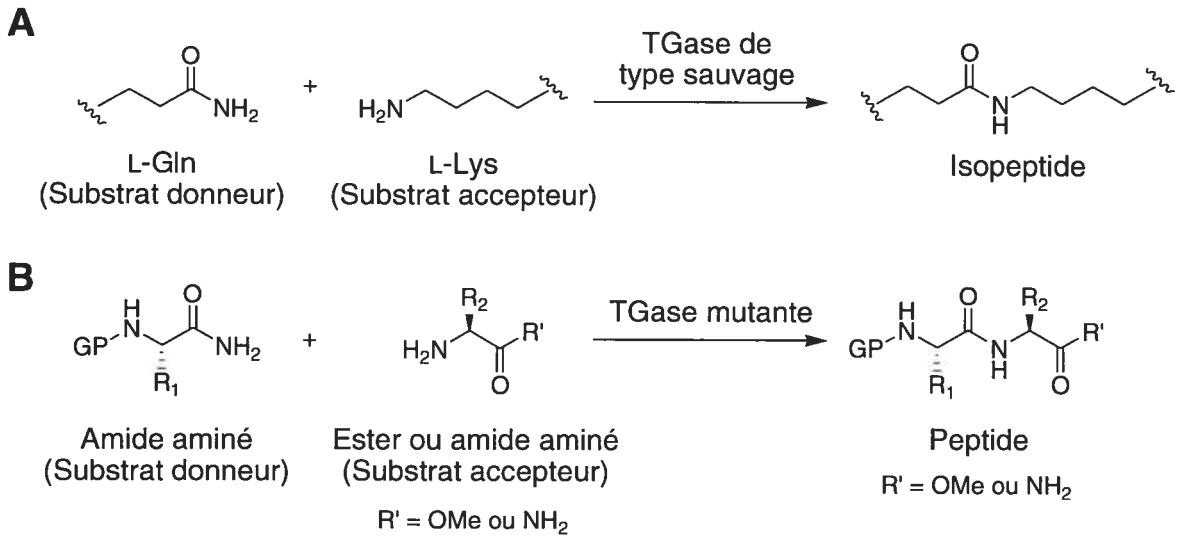


Figure 7. Synthèse des peptides catalysée par une TGase modifiée. (A) Réaction native catalysée par la TGase de type sauvage. (B) Réaction désirée catalysée par un mutant de la TGase. GP indique le groupement protecteur.

Pour atteindre notre objectif, nous proposons d'utiliser une TGase tissulaire, soit la TGase de foie de cobaye. L'avantage d'utiliser une TGase tissulaire plutôt qu'une autre

TGase est que cette classe de TGase a été la plus étudiée : sa spécificité envers les substrats donneurs et accepteurs non naturels est bien caractérisée, plusieurs méthodes d'expression et de purification chez la bactérie ont été rapportées, des structures cristallines ont été élucidées et leur mécanisme catalytique est bien établi. Parmi les TGases tissulaires, c'est la TGase de foie de cobaye qui est la plus étudiée et pour cette raison, c'est elle qui est utilisée tout au long de ce projet de recherche.

1.4.2 Méthodologie générale

Pour atteindre notre objectif, nous avons entrepris de modifier la spécificité de la TGase de foie de cobaye. Puisque la TGase catalyse déjà la formation des liens isopeptidiques à partir des amides primaires (substrat donneur) et des amines primaires (substrat accepteur), nous proposons de modifier les sous-sites de liaison de ces substrats sans modifier la réaction catalysée par l'enzyme. La modification du sous-site de liaison du substrat donneur pourra permettre le remplacement du groupe γ-carboxamide de la glutamine du substrat natif par le groupe α-carboxamide du substrat désiré. Ceci nécessitera la création d'espace dans le site de liaison pour accommoder la chaîne latérale du résidu N-terminal du peptide croissant (R_1) (Figure 7). De façon similaire, le site de liaison du substrat accepteur devra être agrandi pour permettre l'utilisation, comme substrat accepteur, des acides aminés avec des chaînes latérales plus volumineuses (R_2).

La modification de la spécificité de la TGase de foie de cobaye peut se faire par évolution dirigée. L'évolution dirigée est le processus de création, dans un court laps de temps, d'un grand nombre de protéines mutantes (une banque) qui sont criblées pour une activité désirée. L'évolution dirigée imite donc les stratégies de l'évolution naturelle, en accéléré [74-78]. Pour entreprendre l'évolution dirigée d'une enzyme de sorte à obtenir des mutants ayant les nouvelles caractéristiques voulues, il faut procéder aux deux étapes suivantes :

1. La création de banques de mutants de la protéine d'intérêt, par des stratégies telles la mutagenèse à haut taux d'erreurs ou encore le *gene shuffling* où les variantes d'une séquence d'ADN sont fragmentées puis les fragments recombinés afin d'en obtenir de nouvelles combinaisons [79, 80].
2. L'identification des mutants ayant la propriété désirée. Ceci peut être fait par criblage, dont les conditions devraient diriger l'évolution vers le but désiré.

La création de banques d'ADN est maintenant aisément praticable dans les laboratoires ayant l'expertise requise. Par contre, la stratégie d'identification des mutants voulus doit être développée conformément à chaque but, de façon à s'assurer que les nouvelles enzymes sélectionnées possèdent les caractéristiques spécifiquement requises.

L'évolution dirigée des enzymes peut se faire par une approche aléatoire ou par une approche semi-aléatoire [81]. L'approche aléatoire consiste à effectuer la mutagenèse aléatoire sur l'ensemble du gène encodant l'enzyme, c'est-à-dire introduire des mutations au hasard dans l'espoir d'en trouver qui donnent l'activité désirée et qui n'auraient pas pu être prédites. L'approche semi-aléatoire consiste à entreprendre une mutagenèse semi-aléatoire, c'est-à-dire introduire des mutations au hasard, mais seulement dans une ou des régions spécifiques du gène tels les résidus du site actif. La mutagenèse semi-aléatoire peut être combinatoire. Cette approche, de plus en plus répandue, a l'avantage d'exploiter les données connues sur la structure et/ou la fonction de certains résidus, générant ainsi des banques « intelligentes » ayant une meilleure probabilité de contenir des mutants aux propriétés désirées. Cette approche semi-aléatoire, qui fait l'objet du chapitre 2, a été choisie comme méthodologie générale pour atteindre notre objectif de recherche.

1.4.3 Étapes du projet

Afin de procéder à l'évolution dirigée de la TGase de foie de cobaye par une aproche semi-aléatoire, il faut posséder de l'information structurelle et fonctionnelle ainsi que des méthodes de mutagenèse semi-aléatoire et de criblage. L'information structurelle et

fonctionnelle a dû être acquise lors de ce projet de recherche, et les méthodes de mutagenèse semi-aléatoire et de criblage ont dû être développées. Les étapes du projet de recherche ont donc été les suivantes :

1. Identification des résidus de la TGase impliqués dans la liaison du substrat donneur;
2. Développement d'une méthode d'expression et de purification de la TGase de foie de cobaye chez *Escherichia coli*;
3. Développement d'un modèle d'homologie de la TGase de foie de cobaye et mutagenèse de résidus du site actif pour obtenir de l'information structurelle et fonctionnelle supplémentaire;
4. Évolution dirigée de la spécificité de la TGase de foie de cobaye par une approche semi-aléatoire.

Nous décrirons ces étapes dans les sections suivantes.

1.4.3.1 Identification des résidus de la TGase impliqués dans la liaison du substrat donneur

Avant d'entreprendre l'évolution dirigée de la spécificité de la TGase de foie de cobaye par une approche semi-aléatoire, il a fallu identifier les résidus appropriés à muter, c'est-à-dire les résidus formant les sous-sites de liaison des substrats de l'enzyme. Puisque aucune structure cristalline de cette enzyme avec un substrat lié au site actif n'est publiée, nous avons généré un modèle de la TGase de *Pagrus major* avec un substrat donneur lié au site actif. Cette étude a été publiée en 2004 et forme le chapitre 3 de la présente thèse.

1.4.3.2 Développement d'une méthode d'expression et de purification de la TGase de foie de cobaye chez *Escherichia coli*

Après avoir identifié les résidus de la TGase que nous voulions muter, nous avons développé une méthode d'expression et de purification efficace de la TGase de foie de cobaye chez *E. coli* pour obtenir rapidement, efficacement et avec un bon rendement de

l'enzyme soluble ayant une bonne activité spécifique. Ce protocole, développé avec le Dr Steve Gillet, fut publié en 2004 et forme le chapitre 4.

1.4.3.3 Développement d'un modèle d'homologie de la TGase de foie de cobaye et mutagenèse de résidus du site actif

Afin d'obtenir de l'information structurelle et fonctionnelle supplémentaire, nous avons généré un modèle d'homologie de la TGase de foie de cobaye. Ce modèle a permis d'étudier le rôle de certains résidus du site actif. L'information recueillie par cette étude décrite au chapitre 5 nous a servi dans la planification de la mutagenèse semi-aléatoire pour l'évolution dirigée par une approche semi-aléatoire.

1.4.3.4 Évolution dirigée de la spécificité de la TGase de foie de cobaye par une approche semi-aléatoire

À partir de l'information acquise par les études de modélisation des chapitres 3 et 5, nous avons généré cinq banques de mutants de la TGase de foie de cobaye par mutagenèse de saturation et/ou combinatoire. Par la suite, une méthode fluorimétrique de criblage en plaques à 96 puits a été développée afin de permettre l'identification des mutants de ces cinq banques ayant la spécificité désirée. Ces méthodes nous ont permis de procéder à l'évolution dirigée de la spécificité de la TGase de foie de cobaye, qui est décrite au chapitre 6.

Le processus d'évolution dirigée a suivi le cheminement présenté à la Figure 8. Par modélisation moléculaire, nous avons identifié les résidus du site actif qui semblent être impliqués dans la spécificité (étape 1). Ces résidus ont été soumis à une mutagenèse semi-aléatoire (étape 2). Les banques d'ADN ainsi obtenues ont été transformées chez *E. coli* où les TGases mutantes ont été exprimées selon le protocole que nous avons développé (étape 3). Finalement, les banques de TGases mutantes ont été criblées par notre méthode de criblage fluorimétrique (étape 4).

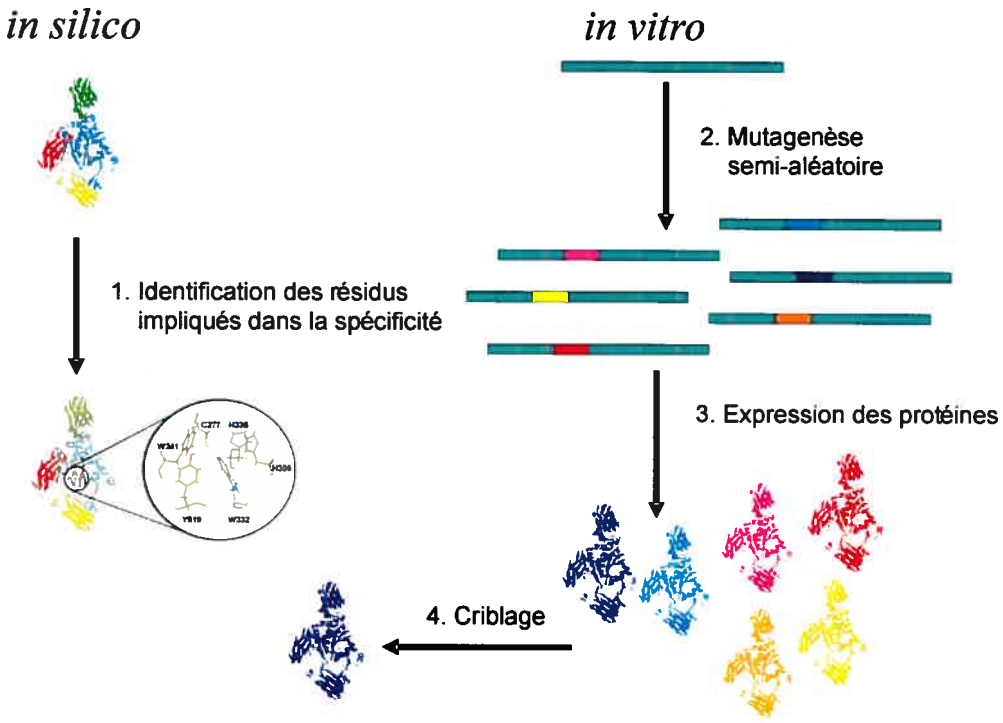


Figure 8. Évolution dirigée de la TGase par une approche semi-aléatoire.

CHAPITRE 2

Évolution dirigée par une approche semi-aléatoire : revue de la littérature

2.0 Préface

Plusieurs méthodes existent pour modifier les propriétés des enzymes selon nos besoins. Pour modifier la spécificité de la TGase de foie de cobaye, nous avons opté pour la méthode de l'évolution dirigée selon une approche semi-aléatoire. Cette approche a été choisie pour deux raisons. Premièrement, nous voulions modifier les sous-sites de liaison des substrats de la TGase sans modifier la réaction catalysée par l'enzyme. En ciblant la mutagenèse au site actif, nous avons l'avantage d'exploiter les données connues sur la structure et/ou la fonction de certains résidus, générant ainsi des banques de mutants « intelligentes ». Ces banques ont une meilleure probabilité de contenir des mutants aux propriétés désirées. Deuxièmement, nous ne possédons pas de méthodes de criblage à haut débit nous permettant de cribler des dizaines de milliers de mutants et plus. L'approche semi-aléatoire semblait donc être le meilleur choix pour atteindre notre objectif.

Dans ce chapitre, la méthode d'évolution dirigée par une approche semi-aléatoire, aussi connue sous le nom d'approche semi-rationnelle, est décrite sous la forme d'un article de revue de la littérature publié en 2004 et intitulé « Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design ». Cet article présente une vue d'ensemble d'articles récents où la modification d'activités enzymatiques par cette approche a été démontrée avec succès. Cet article a été écrit à contribution égale avec Nicolas Doucet. Ma contribution a été de rédiger l'introduction de même que la section sur l'utilisation d'outils informatiques pour l'évolution dirigée par une méthode semi-aléatoire (« Semi-rational and combinatorial design using computational approaches »).

Article 1.

Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design

Roberto A Chica[1,*], Nicolas Doucet[2,*] and Joelle N Pelletier[1,2]

[1]Département de chimie, Université de Montréal, CP 6128, Succursale Centre-Ville, Montréal, Québec, H3C 3J7, Canada

[2]Département de biochimie, Université de Montréal, CP 6128, Succursale Centre-Ville, Montréal, Québec, H3C 3J7, Canada

[*] These authors made an equal contribution to this work.

Current Opinion in Biotechnology, 2005, **16**, 378-384

2.1 Abstract

Many research groups successfully rely on whole-gene random mutagenesis and recombination approaches for the directed evolution of enzymes. Recent advances in enzyme engineering have used a combination of these random methods of directed evolution with elements of rational enzyme modification to successfully by-pass certain limitations of both directed evolution and rational design. Semi-rational approaches that target multiple, specific residues to mutate on the basis of prior structural or functional knowledge create 'smart' libraries that are more likely to yield positive results. Efficient sampling of mutations likely to affect enzyme function has been conducted both experimentally and, on a much greater scale, computationally, with remarkable improvements in substrate selectivity and specificity and in the *de novo* design of enzyme activities within scaffolds of known structure.

2.2 Introduction

Enzymes are powerful catalysts that are able to increase reaction rates by up to 17 orders of magnitude [1 and 2]. Certain enzymes display perfect control over stereochemistry and regioselectivity, while others display a breadth of specificity; either feature may be attractive for industrial and synthetic applications [3]. However, few enzymes naturally catalyze the reactions that chemists need under conditions that are industrially convenient and economically advantageous. Many efforts are currently devoted to the modification of enzyme activities to meet the needs of today's chemists.

Multiple approaches have been developed to allow the identification of mutant enzymes possessing desirable qualities such as increased activity, modified specificity, selectivity or cofactor binding [4]. The earliest approach was rational design, which was used to modify the specificity of enzymes [5, 6, 7, 8 and 9]. This approach requires an in-depth knowledge of the structural features of the enzyme active-site and their contribution

to function. The complexity of the structure/function relationship in enzymes has proven to be the factor limiting the general application of rational design. More recently, directed evolution has proven to be a powerful tool for the modification of enzyme activities and has become the most widely used approach. Briefly, directed evolution consists of the low frequency introduction of randomly distributed mutations in a gene of interest, followed by selection of the mutated proteins possessing the desired properties. Recent advances in this area include the application of the immune system 'hypermutation' process, which is conducted entirely in live cells [10]. Directed evolution can enable the relatively rapid engineering of enzymes without requiring an in-depth understanding of structure/function relationships, unlike rational design. The numerous methodologies and successes of directed evolution will not be detailed here as recent reviews have been devoted to the topic [11, 12, 13, 14 and 15] (also see the article by AA Henry and FE Romesberg in this issue). Because large numbers of mutants must generally be screened to obtain a significant, desired effect on enzyme activity, the main limitation of directed evolution is the necessity of developing a high-throughput screening methodology that allows identification of the desired property under relevant conditions. Not all enzyme activities are readily amenable to developing a high-throughput screening method, nor are all screening methodologies easy to implement at the required scale.

A further limitation of directed evolution relates to the random nature of the mutations introduced: in the case where functional information (from point mutations, random mutagenesis or deduction by sequence alignment) or structural information exists, it would be advantageous to exploit this by concentrating mutations where they might be the most effective [16]. Analysis of enzyme modification results indicates that the majority of mutations that beneficially affect certain enzyme properties (enantioselectivity, substrate specificity and new catalytic activities) are located in or near the active-site, and more specifically near residues that are implicated in binding or catalysis [15, 17, 18• and 19]. Other properties (stability and activity) can be improved by mutations either near or far from the active-site, although there are many more mutations far from the active-site to test

out [19]. Consequently, saturation mutagenesis (where all 20 natural amino acids are tested at residues in or near the active-site) can be coupled to recombination of these mutations in order to increase the likelihood of beneficially modifying the catalytic activity relative to random mutagenesis approaches. This 'semi-rational' approach could be particularly advantageous in instances where no high-throughput screening method is available, based on the argument that a 'smarter' library can be built with the same number of mutants than with a random whole-gene mutagenesis approach [15] (Table 1; Figure 1).

Table 1. Comparison of approaches for engineering enzyme activity

	Rational Design	Random Mutagenesis	Semi-rational Design
High-throughput screening or selection method	Not essential	Essential	Advantageous but not essential
Structural and/or functional information	Both essential	Neither essential	Either is sufficient
Sequence space exploration	Low	Moderate, random	Experimental: moderate, targeted; Computational: vast, targeted
Probability of finding coupled mutations	Moderate	Low	High

This review provides an overview of recent articles where the semi-rational design of enzyme activities has been successfully demonstrated through the use of combinatorial

mutagenesis biased toward the active-site, based both on experimental and computational approaches.

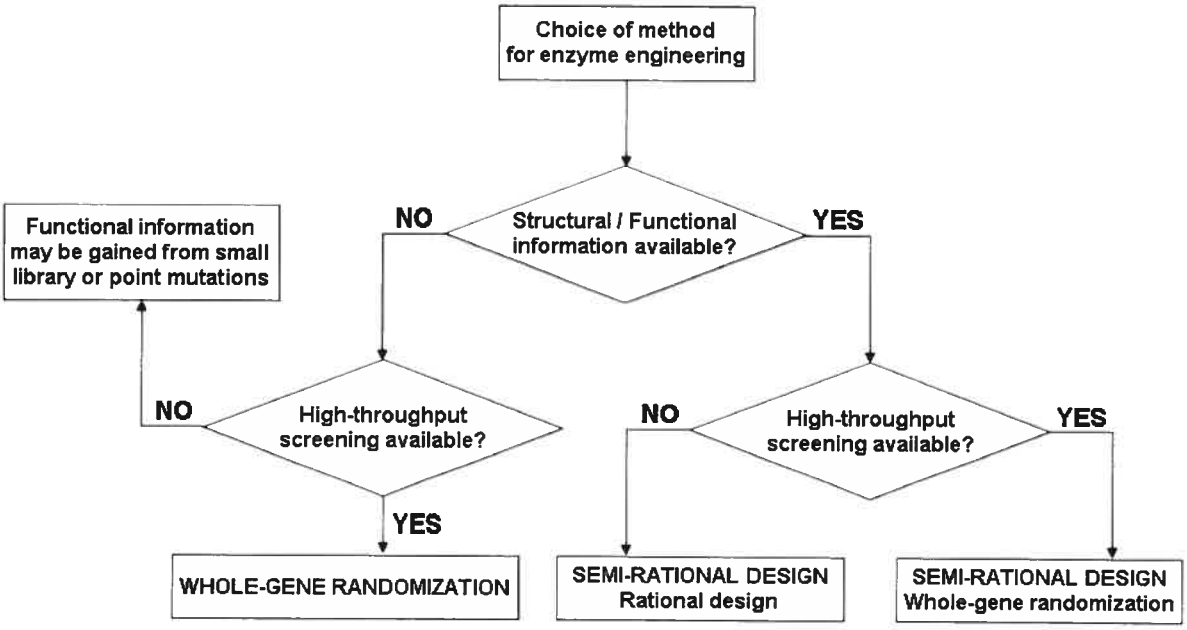


Figure 1. Selection of the preferred experimental approach for enzyme engineering based on the availability of experimental tools and prior knowledge of structure and function. Rational design, semi-rational design or whole-gene randomization each refer to multiple methodologies, as outlined in the text. The enzyme engineering approach that may have the greatest potential for success is in upper case letters, while alternative approaches are in lower case letters.

2.3 Modifying enzyme activity using experimental semi-rational approaches

The experimental combination of rational and random protein engineering approaches has been successfully applied toward the modification of enzyme activities. The

following examples have been grouped according to the specific experimental approaches applied, which are guided by the amount and nature of structural and functional information available at the outset of the study.

2.3.1 Targeted randomization of defined residues based on structural knowledge

Structural information, when available, can allow one to target specific residues in direct contact with the substrate or near the active-site cavity for mutagenesis, alone or in combination. Although the choice of the positions to mutate remains rational in most cases, the choice of amino acids to be encoded can be broad; furthermore, the simultaneous randomization at targeted positions may result in synergistic effects [20] that could not have been predicted by mutating positions individually.

Santoro and Schultz [21•] have used targeted randomization to modify the substrate specificity of a Cre DNA recombinase from bacteriophage P1 that recognizes DNA sequences known as *loxP* sites. The researchers used structural information of the Cre–DNA complex to target two regions of the Cre recombinase in contact with *loxP* base pairs suspected to be important for recombination activity. By creating two distinct libraries (C1 and C2; see Figure 2a) of five and six simultaneously randomized residues in contact with variants of *loxP* sites (library size ~10^8) and by subjecting the *E. coli*-transformed libraries to a powerful recombination assay of fluorescent reporter proteins sorted by FACS (fluorescence-activated cell sorting), they isolated a mutant that efficiently recombined a new *loxP* site not recognized by the wild-type enzyme. Moreover, this Cre mutant retained the ability to recognize the native *loxP* site. The authors favored this targeted library approach because they anticipated that a higher likelihood of success would be observed, while requiring fewer rounds of selective amplification and mutagenesis compared with random mutagenesis of the whole gene.

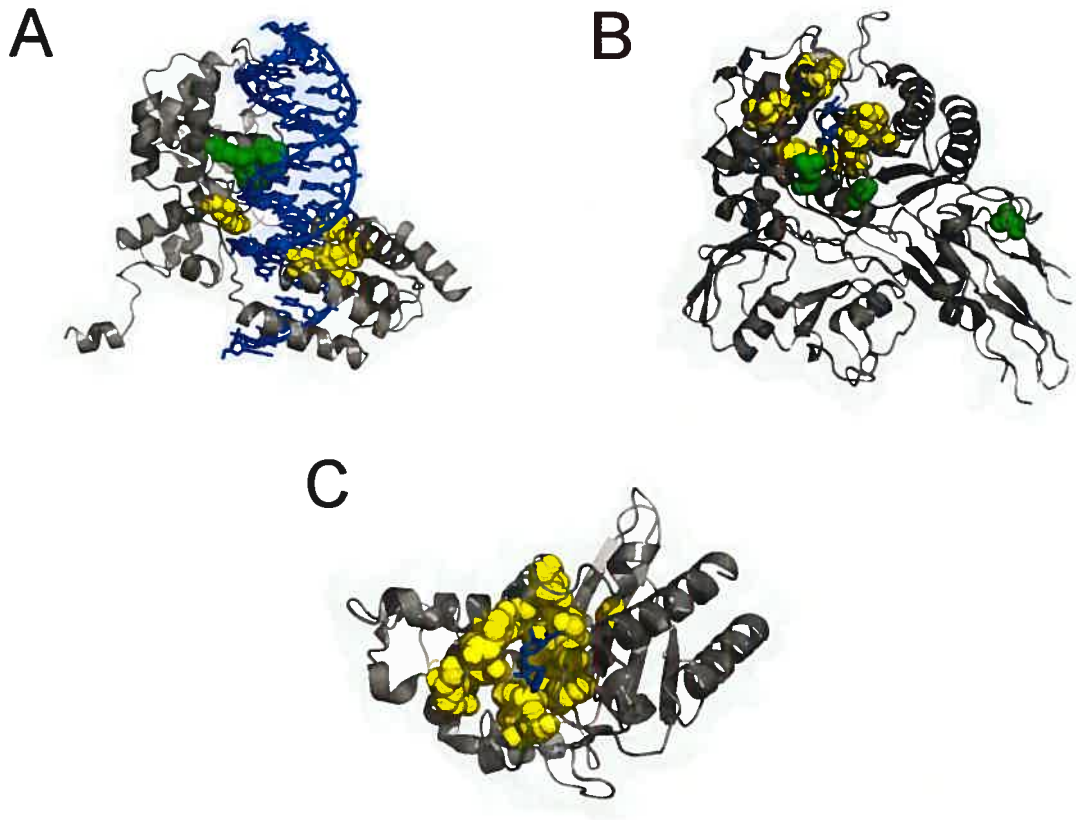


Figure 2. Residues in or near the enzyme active-site that were targeted for semi-rational combinatorial mutagenesis. The enzymes are shown as grey ribbon diagrams with space-filling in semi-transparent gray. (A) Cre DNA recombinase from bacteriophage P1 [21] (PDB accession number 1CRX). Residues from libraries C1 (six residues) and C2 (five residues) and the DNA substrate are colored yellow, green and blue, respectively. (B) PBP2X DD-transpeptidase from *S. pneumonia* [30] (PDB accession number 1QMF). The ten residues targeted for combinatorial site-directed mutagenesis are colored yellow and the three subsequent unforeseen beneficial mutations obtained by epPCR are colored green. Cefuroxime is acylated at the active-site residue Ser337 and is colored blue. (C) TEM-1 β-lactamase from *E. coli* [43] (PDB accession number 1FQG). The 19 active-site residues targeted for *in silico* mutagenesis using PDA are colored yellow. The benzylpenicillin antibiotic is acylated at the active-site residue Ser70 and is colored blue. Figure generated with Pymol Release 0.98 (http://www.pymol.org).

Similarly, in a successful attempt to investigate the detoxification role of epoxide hydrolases, Rui *et al.* [22] used a combination of rational design and saturation mutagenesis at targeted active-site positions to expand the substrate range of an epoxide hydrolase from *Agrobacterium radiobacter* AD1 (EchA) to include chlorinated epoxides. On the basis of a careful investigation of the active-site by structural comparison to related enzymes, residues F108, C248, I219 and I111 (in single-letter amino acid code), in the vicinity of the catalytic triad, were separately randomized by saturation mutagenesis. As mutant C248I showed a slight increase in rate of *cis*-1,2-dichloroethylene (*cis*-DCE) mineralization (2.7-fold increase with respect to the wild-type), it was used as a template for successive rounds of saturation mutagenesis targeting residues F108 and I219. The approach generated the triple active-site area mutant F108L-I219L-C248L, which displayed a 10-fold enhancement in *cis*-DCE mineralization relative to wild-type EchA.

Several other groups [23, 24, 25 and 26] have also successfully conducted simultaneous or successive targeted randomization of multiple active-site positions based on structural knowledge, as an efficient means to circumvent limitations inherent either to site-directed mutagenesis or to whole-gene random mutagenesis. Recently, Schultz and colleagues [27] combinatorially randomized two active-site residues of an aminoacyl-tRNA synthetase as a step in the generation of an orthogonal synthetase/tRNA pair that efficiently and selectively incorporates an unnatural amino acid into proteins.

In some cases, the structural information available is not sufficient to rationally select residues to be randomized. Thus, molecular modeling studies have been successfully used to identify residues most likely to be in contact with substrate molecules. Our group used this approach to perform combinatorial targeted mutagenesis of all 16 principal active-site residues of type II R67 dihydrofolate reductase, allowing the selection of new, highly modified active-site environments with activity at least as great as the native enzyme [28]. Similarly, the substrate specificity of *Bacillus stearothermophilus* SD1 D-hydantoinase was

modified by a semi-rational mutagenesis approach targeting residues chosen on the basis of previous modeling studies [29].

2.3.2 Simultaneous random mutagenesis and site-saturation of defined residues

Based on evolutionary protein-fold similarity between two related enzyme families, Peimbert and Segovia [30] introduced a β-lactamase activity into PBP2X DD-transpeptidase from *Streptococcus pneumonia* through a combination of error-prone polymerase chain reaction (epPCR) and saturation mutagenesis at targeted positions. This approach was used to increase the odds of obtaining beneficial mutations at unpredictable locations in combination with mutations at carefully chosen residues. This approach should increase the return using the same number of mutants as a random whole-gene mutagenesis scheme. Thus, two active-site residues (F450 and W374) were targeted for saturation mutagenesis in combination with simultaneous amino acid replacements at eight positions near the active-site cavity (see Figure 2b). After the oligonucleotide-based mutagenesis, the gene was further amplified with a 0.8 % random mutagenesis rate. This library was selected for cefotaxime resistance and mutants with 10-fold increased resistance relative to wild-type PBP2X were obtained. The random mutagenesis resulted in additional mutations at positions 312, 452 and 554, both proximal to and distal from the active-site. To assess the impact of certain randomly inserted mutations, saturation mutagenesis was undertaken at positions 312, 336, 450 and 452. Although no further increase in cefotaxime resistance was observed compared with the original mutants, one new mutant showed β-lactamase activity without compromising the original DD-peptidase activity of PBP2X, thus generating a mutant with dual substrate specificity.

2.3.3 Random mutagenesis followed by site-saturation of defined residues

The structural or functional information necessary to make rational choices for the residues to mutate is not always available. To circumvent this limitation, several research groups have efficiently undertaken rounds of whole-gene randomization to provide 'leads' (i.e. residues identified as potentially advantageous when mutated) in conjunction with fine-tuning of the 'lead mutants' through site-saturation mutagenesis of the identified residues. Using such a cyclical random/targeted approach, Reetz and collaborators greatly improved the enantioselectivity *(E)* of *Pseudomonas aeruginosa* lipases toward a *p*-nitrophenyl ester in favour of the (2*S*) enantiomer (reviewed in [31••]). Initial improvement from $E = 1.1$ (wild-type enzyme) to $E = 11.3$ was achieved after four rounds of epPCR.

Using the mutational knowledge obtained by epPCR, they conducted saturation mutagenesis at given mutational 'hot spots' of the lipase, obtaining $E = 20$. Subsequent rounds of epPCR increased E to 25. In addition, they developed a modified version of Stemmer's combinatorial multiple-cassette mutagenesis method [32] to mutate a defined region of the lipase (residues 160–163) as well as several 'hot spots' (residues 155 and 162). This semi-rational method yielded their most highly enantioselective lipase variant to date ($E = 51$). Other groups have recently demonstrated the power of similar semi-rational approaches toward the evolution of enantioselective enzymes. As a result of the multiple random mutagenesis steps, this approach has also generated many mutations far from the active-site cavity in addition to mutations designed within the active-site cavity [33, 34 and 35].

Geddie and Matsumura [36] have modified the substrate specificity of *Escherichia coli* β-glucuronidase through combinatorial site-saturation mutagenesis of several 'hot spots', using knowledge previously obtained from DNA shuffling studies. The resulting mutants displayed up to a 70-fold increase in xylosidase activity and were further improved

by several steps of whole-gene random mutagenesis using DNA shuffling, epPCR and Staggered Extension Process (StEP) to generate mutants displaying 100-fold improvement in xylosidase activity compared with their ancestor. Similar studies were undertaken to modify the enzymatic specificity and activity toward 7-aminodesacetoxycephalosporanic acid (7-ADCA) of a glutaryl acylase of *Pseudomonas* SY-77 [37] and to modify and improve fluorescent proteins from *Discosoma* sp. [38•]. In all of these studies, whole-gene randomization was generally undertaken in a first round of mutagenesis, followed by 'fine-tuning' of the most interesting mutants through single or combinatorial site-saturation mutagenesis at targeted positions. Afterwards, mutants obtained were further improved using subsequent rounds of random mutagenesis.

In the absence of structural or functional information, the 'Evolutionary Trace' method [39] can pinpoint residues to mutate. This method entails correlating evolutionary variations within a gene of interest with divergences in the phylogenetic tree of that sequence family. This has been shown to reveal the relative functional importance of residues and to identify functional sites [40]. A recent review by Minshull *et al.* [41] provides a highly comprehensive overview of this and other approaches for identifying residues that may be functionally relevant.

These recent examples demonstrate the power of structure-based semi-rational approaches targeting active-site residues in enzyme engineering by enhancing our capacity for rational design, while exploiting the advantageous aspects of random mutagenesis. We can thus cycle between light sampling of randomly distributed mutations and saturation mutagenesis at a limited number of positions likely to affect the property under study, increasing the likelihood of identifying beneficial, cooperative effects with respect to enzyme catalysis.

2.4 Semi-rational and combinatorial design using computational approaches

Although semi-rational, combinatorial mutagenesis directed at active-site residues limits the number of variants generated relative to available sequence space and theoretically focuses on an area of sequence space where improvements have a greater likelihood of occurring, it can nonetheless generate greater numbers of mutants than can be screened. Indeed, combinatorial randomization of only five residues generates a library of 20^5 possibilities (3.2×10^6 mutants), too large a number for manual screening. Thus, to increase the power of semi-rational and combinatorial modification of enzyme activities, computational methods have been developed based on protein design algorithms. These methods can either perform a virtual screening of a vast library or can be applied to the design of enzyme active-sites.

The first method is the computational screening of mutant sequences of a virtual library. Computational screening can screen libraries of 10^{80} variants [42], allowing for a first layer of virtual screening to eliminate mutations inconsistent with the protein fold. This is an emerging area in protein design and has seldom yet been applied specifically to enzyme design. However, its resounding success and great promise merit discussion herein. Hayes *et al.* [43•] developed a strategy for the computational screening of large libraries called Protein Design Automation (PDA), for prediction of the optimal sequence that can adopt a desired fold. PDA was used for prescreening large, virtual libraries of mutants (10^{23}), thus decreasing the sequence space of interest by many orders of magnitude. PDA allows all, or a rationally defined set of residues, to change. The optimal sequence is chosen based on its lowest conformational energy and is used to identify other near-optimal sequences through Monte Carlo simulated annealing. The mutations that occur most frequently define the library to be experimentally screened. Using PDA, Hayes and colleagues pinpointed 19 residues of interest in TEM-1 β-lactamase (see Figure 2c),

generated *in silico* 7×10^{23} combinatorial mutants of these residues, and chose cut-offs to define the library of roughly 200 000 lowest-energy mutants that were then generated experimentally by mutagenesis and recombination. By selection against the antibiotic cefotaxime, the authors identified a mutant harbouring six mutations in the vicinity of the active-site, with a 1280-fold increase in resistance to cefotaxime that was attributed to an increase in relative substrate specificity toward cefotaxime versus ampicillin.

The advantage of PDA is that it samples a vast sequence diversity and allows for multiple mutations to be identified simultaneously, which is particularly beneficial when the effect of multiple mutations is synergistic (non-additive) [20 and 44]. Furthermore, PDA generates mutations at the level of the amino acid sequence rather than at the level of the nucleotide sequence. Thus, there is no bias against mutations requiring two or three nucleotide modifications, contrary to the important bias that exists in standard random mutagenesis methods. This approach vastly increases explorable sequence space, but has not been specifically designed for improvement of enzyme activity.

The second computational method for semi-rational and combinatorial design, which has been specifically applied to the design of enzyme active-sites, was developed by Hellinga and co-workers [45••]. This ground-breaking work, involving the prediction of mutations that are necessary for the introduction of catalytically active-sites in non-catalytic protein scaffolds, allowed the authors to achieve the most dramatic success in computational enzyme design to date. In their report, Dwyer *et al.* [45••] converted a non-catalytic ribose-binding protein (RBP) into an analog of triose phosphate isomerase (TIM). Their protein design algorithm first predicted mutations at the ribose-binding area to allow binding of the TIM substrate, dihydroxyacetone phosphate (DHAP) [46]. The algorithm then positioned a set of TIM catalytic residues in the new DHAP-binding pocket. Finally, 14 virtual constructs were experimentally generated and tested for TIM activity. Seven of the mutants possessed TIM activity greater than the background reaction. One of the seven designs was particularly active and through further improvements by computational design

and random mutagenesis, allowed the generation of TIM analogs from a non-catalytic protein. The computational design allowed for the definition of 13 to 21 mutations, depending on the TIM analog, that introduce the necessary catalytic residues as well as residues forming a stereochemically complementary substrate-binding cavity. The most active TIM analogs displayed a 10^5–10^6-fold increase in reaction rate over background, which is the largest increase in reaction rate for a rationally designed enzyme to date [47, 48 and 49], and was sufficiently active to support growth of TIM-deficient *E. coli*. An important contributing factor to the success of this work results from their superior treatment of electrostatic interactions in a hetereogeneous protein environment [50]. Although this formidable achievement required prior knowledge of the precise chemical and steric requirements of the target activity, it opens the door to the broader creation of desirable catalytic activities within stable and well-behaved frameworks.

2.5 Conclusions

Targeted, combinatorial semi-rational mutagenesis is proving highly effective for improving enzyme activities, as it readily allows the creation of neighbouring mutations, of multiple simultaneous mutations and of mutations requiring multiple nucleotide substitutions. This could be particularly advantageous in the modification of enzyme activities, as active-site mutations are frequently coupled and have synergistic effects [20 and 44]. Important developments in computational methodologies promise to vastly increase the searchable sequence space. Taken with a judicious choice of experimental input, semi-rational mutagenesis permits the experimenter to focus mutations in areas more likely to yield 'lead' results. These methodologies pave the way to exciting areas of enzyme research including efficient modification of existing activities, the development of new activities within existing frameworks, as well as the evolution of 'promiscuous' catalytic activities allowing their efficient exploitation [51 and 52].

2.6 Acknowledgements

The authors thank Romas J Kazlauskas for his careful reading of the manuscript and his helpful comments.

2.7 References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1 A. Radzicka and R. Wolfenden, A proficient enzyme, *Science* **267** (1995), pp. 90–93.

2 S.J. Benkovic and S. Hammes-Schiffer, A perspective on enzyme catalysis, *Science* **301** (2003), pp. 1196–1202.

3 O. Kirk, T.V. Borchert and C.C. Fuglsang, Industrial enzyme applications, *Curr Opin Biotechnol* **13** (2002), pp. 345–351.

4 U.T. Bornscheuer and M. Pohl, Improved biocatalysts by directed evolution and rational protein design, *Curr Opin Chem Biol* **5** (2001), pp. 137–143.

5 N.S. Scrutton, A. Berry and R.N. Perham, Redesign of the coenzyme specificity of a dehydrogenase by protein engineering, *Nature* **343** (1990), pp. 38–43.

6 C.S. Craik, C. Largman, T. Fletcher, S. Roczniak, P.J. Barr, R. Fletterick and W.J. Rutter, Redesigning trypsin: alteration of substrate specificity, *Science* **228** (1985), pp. 291–297.

7 P. Carter, B. Nilsson, J.P. Burnier, D. Burdick and J.A. Wells, Engineering subtilisin BPN' for site-specific proteolysis, *Proteins* **6** (1989), pp. 240–248.

8 J.A. Wells, D.B. Powers, R.R. Bott, T.P. Graycar and D.A. Estell, Designing substrate specificity by protein engineering of electrostatic interactions, *Proc Natl Acad Sci USA* **84** (1987), pp. 1219–1223.

9 F. Cedrone, A. Ménez and E. Quéméneur, Tailoring new enzyme functions by rational redesign, *Curr Opin Struct Biol* **10** (2000), pp. 405–410.

10 L. Wang, W.C. Jackson, P.A. Steinbach and R.Y. Tsien, Evolution of new nonantibody proteins via iterative somatic hypermutation, *Proc Natl Acad Sci USA* **101** (2004), pp. 16745–16749.

11 K.E. Jaeger and T. Eggert, Enantioselective biocatalysis optimized by directed evolution, *Curr Opin Biotechnol* **15** (2004), pp. 305–313.

12 J.L. Jestin and P.A. Kaminski, Directed enzyme evolution and selections for catalysis based on product formation, *J Biotechnol* **113** (2004), pp. 85–103.

13 H. Tao and V.W. Cornish, Milestones in directed enzyme evolution, *Curr Opin Chem Biol* **6** (2002), pp. 858–864.

14 G.J. Williams, A.S. Nelson and A. Berry, Directed evolution of enzymes for biocatalysis and the life sciences, *Cell Mol Life Sci* **61** (2004), pp. 3034–3046.

15 P.A. Dalby, Optimising enzyme function by directed evolution, *Curr Opin Struct Biol* **13** (2003), pp. 500–505.

16 K. Chockalingam, Z. Chen, J.A. Katzenellenbogen and H. Zhao, Directed evolution of specific receptor–ligand pairs for use in the creation of gene switches, *Proc Natl Acad Sci USA* **102** (2005), pp. 5691–5696.

17 S. Park, K.L. Morley, G.P. Horsman, M. Holmquist, K. Hult and R.J. Kazlauskas, Focusing mutations into the *P. fluorescens* esterase binding site increases enantioselectivity more effectively than distant mutations, *Chem Biol* **12** (2005), pp. 45–54.

18• S.L. Strausberg, B. Ruan, K.E. Fisher, P.A. Alexander and P.N. Bryan, Directed coevolution of stability and catalytic activity in calcium-free subtilisin, *Biochemistry* **44** (2005), pp. 3272–3279.

A great example of semi-rational design of enzymatic activity and stability based on structural and functional data previously obtained with subtilisin. Jointly rational and semi-rational, the approach is based on a stepwise increase in the activity of mutants obtained by a randomization of targeted residues near the active-site cavity.

19 K.L. Morley and R.J. Kazlauskas, Improving enzyme properties: when are closer mutations better?, *Trends Biotechnol* **23** (2005), pp. 231–237.

20 A.S. Mildvan, Inverse thinking about double mutants of enzymes, *Biochemistry* **43** (2004), pp. 14517–14520.

21• S.W. Santoro and P.G. Schultz, Directed evolution of the site specificity of Cre recombinase, *Proc Natl Acad Sci USA* **99** (2002), pp. 4185–4190.

Using structural knowledge, residues of Cre recombinase from bacteriophage P1 that are in direct contact with variants of the *loxP* DNA substrate were subjected to a semi-rational saturation mutagenesis approach to successfully switch the substrate specificity of the Cre enzyme. A powerful screening method using FACS is also described.

22 L. Rui, L. Cao, W. Chen, K.F. Reardon and T.K. Wood, Active-site engineering of the epoxide hydrolase from *Agrobacterium radiobacter* AD1 to enhance aerobic mineralization of *cis*-1,2-dichloroethylene in cells expressing an evolved toluene ortho-monooxygenase, *J Biol Chem* **279** (2004), pp. 46810–46817.

23 C.M. Hill, W.S. Li, J.B. Thoden, H.M. Holden and F.M. Raushel, Enhanced degradation of chemical warfare agents through molecular engineering of the phosphotriesterase active-site, *J Am Chem Soc* **125** (2003), pp. 8990–8991.

24 M. Wilming, A. Iffland, P. Tafelmeyer, C. Arrivoli, C. Saudan and K. Johnsson, Examining reactivity and specificity of cytochrome *c* peroxidase by using combinatorial mutagenesis, *ChemBioChem* **3** (2002), pp. 1097–1104.

25 N.M. Antikainen, P.J. Hergenrother, M.M. Harris, W. Corbett and S.F. Martin, Altering substrate specificity of phosphatidylcholine-preferring phospholipase C of *Bacillus*

cereus by random mutagenesis of the headgroup binding site, *Biochemistry* **42** (2003), pp. 1603–1610.

26 W.S. Yew, J. Akana, E.L. Wise, I. Rayment and J.A. Gerlt, Evolution of enzymatic activities in the orotidine 5′-monophosphate decarboxylase suprafamily: enhancing the promiscuous D-arabino-hex-3-ulose 6-phosphate synthase reaction catalyzed by 3-keto- L-gulonate 6-phosphate decarboxylase, *Biochemistry* **44** (2005), pp. 1807–1815.

27 J.C. Anderson, N. Wu, S.W. Santoro, V. Lakshman, D.S. King and P.G. Schultz, An expanded genetic code with a functional quadruplet codon, *Proc Natl Acad Sci USA* **101** (2004), pp. 7566–7571.

28 A.R. Schmitzer, F. Lépine and J.N. Pelletier, Combinatorial exploration of the catalytic site of a drug-resistant dihydrofolate reductase: creating alternative functional configurations, *Protein Eng Des Sel* **17** (2004), pp. 809–819.

29 Y.H. Cheon, H.S. Park, J.H. Kim, Y. Kim and H.S. Kim, Manipulation of the active-site loops of D-hydantoinase, a $(\beta/\alpha)_8$-barrel protein, for modulation of the substrate specificity, *Biochemistry* **43** (2004), pp. 7413–7420.

30 M. Peimbert and L. Segovia, Evolutionary engineering of a β-lactamase activity on a D-Ala D-Ala transpeptidase fold, *Protein Eng* **16** (2003), pp. 27–35.

31•• M.T. Reetz, Controlling the enantioselectivity of enzymes by directed evolution: practical and theoretical ramifications, *Proc Natl Acad Sci USA* **101** (2004), pp. 5716–5722.

An impressive example of enzyme evolution using a combination of semi-rational and random mutagenesis approaches to overcome the respective limitations of each method. The author consolidates the progression in modification of enantioselectivity in *P. aeruginosa* lipases from E = 1.1 (wild-type) to E = 51, clearly describing improvements obtained with each mutational step.

32 M.T. Reetz, S. Wilensek, D.X. Zha and K.E. Jaeger, Directed evolution of an enantioselective enzyme through combinatorial multiple-cassette mutagenesis, *Angew Chem Int Ed Engl* **40** (2001), pp. 3589–3591.

33 Y. Koga, K. Kato, H. Nakano and T. Yamane, Inverting enantioselectivity of *Burkholderia cepacia* KWI-56 lipase by combinatorial mutation and high-throughput screening using single-molecule PCR and *in vitro* expression, *J Mol Biol* **331** (2003), pp. 585–592.

34 G.P. Horsman, A.M.F. Liu, E. Henke, U.T. Bornscheuer and R.J. Kazlauskas, Mutations in distant residues moderately increase the enantioselectivity of *Pseudomonas fluorescens* esterase towards methyl 3-bromo-2-methylpropanoate and ethyl 3-phenylbutyrate, *Chemistry* **9** (2003), pp. 1933–1939.

35 B. Lingen, J. Grotzinger, D. Kolter, M.R. Kula and M. Pohl, Improving the carboligase activity of benzoylformate decarboxylase from *Pseudomonas putida* by a combination of directed evolution and site-directed mutagenesis, *Protein Eng* **15** (2002), pp. 585–593.

36 M.L. Geddie and I. Matsumura, Rapid evolution of β-glucuronidase specificity by saturation mutagenesis of an active-site loop, *J Biol Chem* **279** (2004), pp. 26462–26468.

37 C.F. Sio, A.M. Riemens, J.M. van der Laan, R.M. Verhaert and W.J. Quax, Directed evolution of a glutaryl acylase into an adipyl acylase, *Eur J Biochem* **269** (2002), pp. 4495–4504.

38• N.C. Shaner, R.E. Campbell, P.A. Steinbach, B.N. Giepmans, A.E. Palmer and R.Y. Tsien, Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein, *Nat Biotechnol* **22** (2004), pp. 1567–1572.

Simultaneous saturation mutagenesis at many targeted positions coupled to epPCR allowed the authors to generate an impressive array of multiple fluorescent and stable monomer variants of *Discosoma* sp. fluorescent protein (DsRed) based solely on sequence

comparison of mutants of interest. These stable monomers are now applicable to discrimination of cell types, transcriptional activities and/or fusion proteins. Although not oriented toward catalysis, this example nevertheless demonstrates the power of semi-rational approaches.

39 O. Lichtarge, H.R. Bourne and F.E. Cohen, An evolutionary trace method defines binding surfaces common to protein families, *J Mol Biol* **257** (1996), pp. 342–358.

40 O. Lichtarge, H. Yao, D.M. Kristensen, S. Madabushi and I. Mihalek, Accurate and scalable identification of functional sites by evolutionary tracing, *J Struct Funct Genomics* **4** (2003), pp. 159–166.

41 J. minshull, J.E. Ness, C. Gustafsson and S. Govindarajan, Predicting enzyme function from protein sequence, *Curr Opin Chem Biol* **9** (2005), pp. 202–209.

42 B.I. Dahiyat, *In silico* design for protein stabilization, *Curr Opin Biotechnol* **10** (1999), pp. 387–390.

43• R.J. Hayes, J. Bentzien, M.L. Ary, M.Y. Hwang, J.M. Jacinto, J. Vielmetter, A. Kundu and B.I. Dahiyat, Combining computational and experimental screening for rapid optimization of protein properties, *Proc Natl Acad Sci USA* **99** (2002), pp. 15926–15931.

An innovative approach to the use of computational tools in protein engineering. The prescreen described here could be invaluable for the application of a semi-rational approach to enzyme improvement.

44 C.A. Voigt, S. Kauffman and Z.G. Wang, Rational evolutionary design: the theory of *in vitro* protein evolution, *Adv Protein Chem* **55** (2000), pp. 79–160.

45•• M.A. Dwyer, L.L. Looger and H.W. Hellinga, Computational design of a biologically active enzyme, *Science* **304** (2004), pp. 1967–1971.

Computational design of an enzymatic activity in a protein scaffold of known structure. Ground-breaking in that it demonstrates the feasibility of creating new enzymatic activities where there was none by introducing mutations at or near the substrate-binding site.

46 L.L. Looger, M.A. Dwyer, J.J. Smith and H.W. Hellinga, Computational design of receptor and sensor proteins with novel functions, *Nature* **423** (2003), pp. 185–190.

47 D. Hilvert, Critical analysis of antibody catalysis, *Annu Rev Biochem* **69** (2000), pp. 751–793.

48 D.N. Bolon and S.L. Mayo, Enzyme-like proteins by computational design, *Proc Natl Acad Sci USA* **98** (2001), pp. 14274–14279.

49 D.N. Bolon, C.A. Voigt and S.L. Mayo, *De novo* design of biocatalysts, *Curr Opin Chem Biol* **6** (2002), pp. 125–129.

50 M.S. Wisz and H.W. Hellinga, An empirical model for electrostatic interactions in proteins incorporating multiple geometry-dependent dielectric constants, *Proteins* **51** (2003), pp. 360–377.

51 R.J. Kazlauskas, Enhancing catalytic promiscuity for biocatalysis, *Curr Opin Chem Biol* **9** (2005), pp. 195–201.

52 A. Aharoni, L. Gaidukov, O. Khersonsky, Q.G.S. Mc, C. Roodveldt and D.S. Tawfik, The 'evolvability' of promiscuous protein functions, *Nat Genet* **37** (2005), pp. 73–76.

CHAPITRE 3

Identification des résidus de la TGase impliqués dans la liaison du substrat donneur

3.0 Préface

Pour pouvoir entreprendre l'évolution dirigée d'une enzyme selon une approche semi-aléatoire, il faut posséder de l'information structurelle et fonctionnelle. Pour notre projet, l'information structurelle dont nous avions besoin était de connaître le mode de liaison d'un substrat dans le site actif de la TGase. Puisque aucune structure cristalline de TGase avec un substrat lié au site actif n'a été publiée, nous avons généré un modèle moléculaire de ce complexe enzyme/substrat. Ce modèle nous a permis d'identifier les résidus du site actif qui semblent être impliqués dans la liaison du substrat donneur et de proposer des interactions importantes entre l'enzyme et son substrat lors de l'étape d'acylation du mécanisme.

Cette étude de modélisation présentée au chapitre 3 a été publiée en 2004. Pour les travaux qui ont mené à la rédaction du manuscrit « Tissue transglutaminase acylation: proposed role of conserved active-site Tyr and Trp residues revealed by molecular modeling of peptide substrate binding », j'ai effectué toutes les études de modélisation moléculaire assistée par ordinateur à partir d'un protocole que j'ai mis au point, en plus de rédiger le manuscrit. La contribution de Paul Gagnon a été de synthétiser et de caractériser par cinétique enzymatique divers peptides agissant comme substrats donneurs de la TGase de foie de cobaye.

Article 2.

Tissue transglutaminase acylation: Proposed role of conserved active-site Tyr and Trp residues revealed by molecular modeling of peptide substrate binding

Roberto A. Chica, Paul Gagnon, Jeffrey W. Keillor and Joelle N. Pelletier

Département de chimie, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal, Québec, Canada, H3C 3J7

Protein Science, 2004, **13**, 979-991

3.1 Abstract

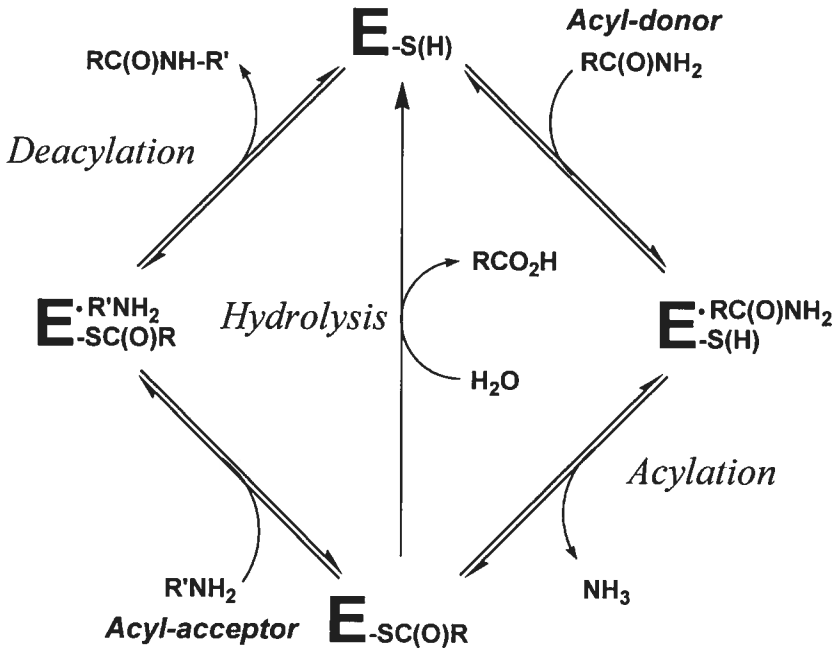
Transglutaminases (TGases) catalyze the cross-linking of peptides and proteins by the formation of γ-glutamyl-ε-lysyl bonds. Given the implication of tissue TGase in various physiological disorders, development of specific tissue TGase inhibitors is of current interest. To aid in the design of peptide-based inhibitors, a better understanding of the mode of binding of model peptide substrates to the enzyme is required. Using a combined kinetic/molecular modeling approach, we have generated a model for the binding of small acyl-donor peptide substrates to tissue TGase from red sea bream. Kinetic analysis of various N-terminally derivatized Gln-Xaa peptides has demonstrated that many Cbz-Gln-Xaa peptides are typical *in vitro* substrates with K_M values between 1.9 mM and 9.4 mM, whereas Boc-Gln-Gly is not a substrate, demonstrating the importance of the Cbz group for recognition. Our binding model of Cbz-Gln-Gly on tissue TGase has allowed us to propose the following steps in the acylation of tissue TGase. First, the active-site is opened by displacement of conserved W329. Second, the substrate Gln side chain enters the active-site and is stabilized by hydrophobic interaction with conserved residue W236. Third, a hydrogen bond network is formed between the substrate Gln side chain and conserved residues Y515 and the acid-base catalyst H332 that helps to orient and activate the γ-carboxamide group for nucleophilic attack by the catalytic sulphur atom. Finally, an H-bond with Y515 stabilizes the oxyanion formed during the reaction.

Keywords: tissue transglutaminase; molecular modeling; docking; enzyme kinetics; peptide substrate; autodock; TGase

3.2 Introduction

Transglutaminases (TGases, EC 2.3.2.13) are enzymes that catalyze the cross-linking of peptides and proteins by the formation of isopeptide bonds between the γ-carboxamide group of a glutamine side chain and the ε-amino group of a lysine side

chain. This reaction is known to occur via a modified ping-pong mechanism (Folk 1969) in which a glutamine-containing protein or peptide, the acyl-donor substrate, reacts with the enzyme's catalytic cysteine residue to form a thioester bond which generates the covalent acyl-enzyme intermediate with concomitant release of ammonia (Scheme 1). This intermediate then reacts with a second substrate, the acyl-acceptor, which can be almost any primary amine (Aeschlimann and Paulsson 1994), to yield the amide product and free enzyme. The acyl-enzyme intermediate can also be hydrolyzed in the absence of primary amine but at a slower rate. A conserved Cys-His-Asp catalytic triad that is similar to that of cysteine proteases catalyzes these reactions. Although TGases exhibit high specificity towards the side chain of L-Gln as the acyl-donor substrate (Asn and D-Gln are not recognized), their specificity towards the acyl-acceptor is lower, and many primary amines can be recognized (Clarke *et al.* 1959; Folk 1983).



Scheme 1. Catalytic mechanism of tissue TGase.

TGases are divided into nine classes (Chen and Mehta 1999), of which the most abundant class is the ubiquitous tissue transglutaminase (Fesus and Piacentini 2002) found in all vertebrates (Wada *et al.* 2002). Tissue TGases are intracellular, monomeric enzymes that exhibit a calcium-dependent transglutaminase activity that is inhibited by GDP/GTP binding. Their structure comprises four domains (Figure 1): the N-terminal β-sandwich domain, the catalytic core, the barrel 1 domain, and the C-terminal barrel 2 domain. Two crystal structures of tissue transglutaminase have been deposited in the Brookhaven Protein Data Bank. These are the structures of red sea bream tissue TGase (1G0D; Noguchi *et al.* 2001) and of human tissue transglutaminase with GDP bound at its allosteric binding site (1KV3; Liu *et al.* 2002). However, no TGase has yet been cocrystallized with ligand bound at the active-site, limiting our understanding of the mode of substrate binding, although kinetic experiments have been performed that give clues as to the structural requirements for acyl-donor and acyl-acceptor substrates of tissue TGase (Clarke *et al.* 1959; Folk and Chung 1973). Folk and Cole (1965) made numerous observations with respect to the structural requirements for small peptide acyl-donor substrates of guinea pig liver TGase. They reported that Gln alone does not act as a substrate, nor do the peptides Gln-Gly and Gly-Gln-Gly. However, Cbz-Gln-Gly, Cbz-Gln-Gly ethyl ester, and benzoyl-Gly-Gln-Gly act as substrates. Cbz-Gln shows poor but detectable activity. Folk and Cole concluded that for a glutamine on a small peptide to be recognized by tissue TGase it must be at least the next-to-last residue and the peptide should bear an N-terminal Cbz group. For this reason, the commercially available Cbz-Gln-Gly peptide serves as one of the most common nonproteic acyl-donor substrates for TGases in the literature.

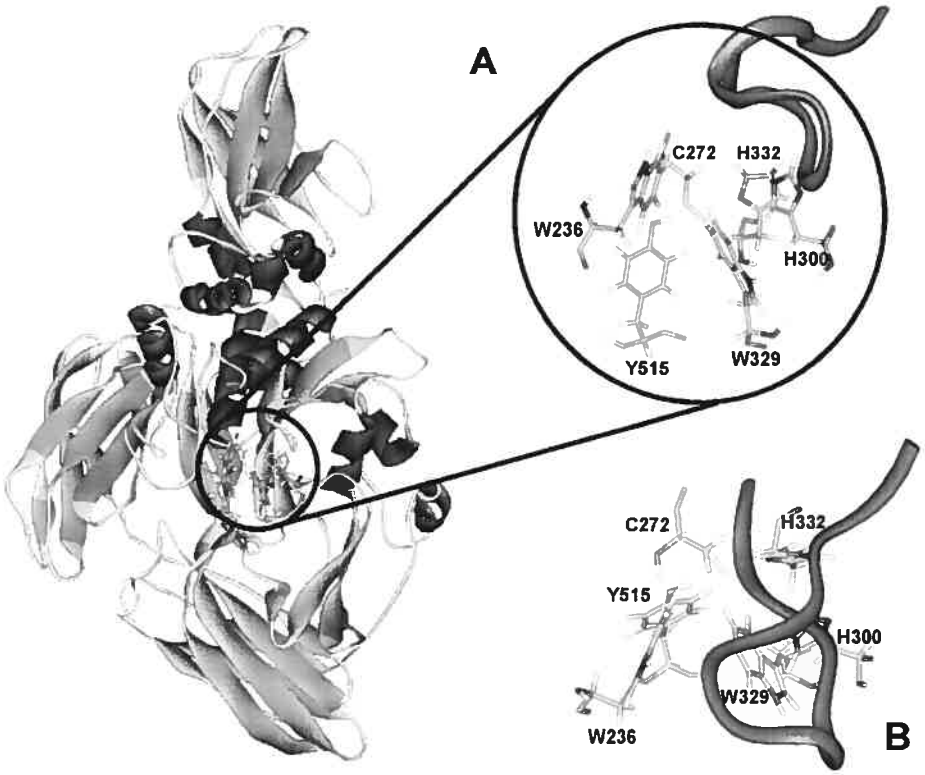


Figure 1. Crystal structure of red sea bream TGase (Noguchi *et al.* 2001). (A) Red sea bream TGase with active-site residues circled. Front view. The P356–G369 loop is represented in ribbon diagram. (B) Top view.

Although there is no consensus as to their physiological role, tissue TGases have been reported to act in endocytosis (Abe *et al.* 2000), apoptosis (Huo *et al.* 2003), extracellular matrix modification (Priglinger *et al.* 2003), and cell signaling (Singh *et al.* 2003). Tissue TGases have also been implicated in various physiological disorders such as Alzheimer's disease (Kim *et al.* 1999), cataract formation (Shridas *et al.* 2001), and more recently, in Celiac sprue (Shan *et al.* 2002). As a result, development of inhibitors specific to tissue TGases (as opposed to type 1 or 3 TGases; Chen and Mehta 1999) is of current interest. However, to aid in inhibitor design, we need to gain a better understanding of the mode of binding of model peptide substrates to the enzyme. To this end, we performed a

combined experimental/molecular modeling strategy consisting of (1) the synthesis of a series of Cbz- or Boc-derivatized peptides and the measurement of their K_M values, and (2) molecular modeling of their binding at the active-site. Correlating the kinetic results with molecular modeling of the ligands bound at the active-site cavity has allowed us to gain a better understanding of the requirements for productive binding of small peptide acyl-donor substrates.

3.3 Results

3.3.1 Kinetic characterization of the N-terminally derivatized peptides

The synthesis of the derivatized peptide substrates has been described (Gagnon *et al.* 2002). The kinetic parameters K_M and k_{cat} of guinea pig liver TGase were determined for a series of Cbz-derivatized Gln-Xaa peptides as well as for the Boc-derivatized Gln-Gly dipeptide (Table 1). The seven Cbz-peptides assayed gave similar K_M and k_{cat} values, whereas the Boc-dipeptide did not appear to act as a donor substrate, showing no reactivity at concentrations up to 40 mM, the limit of its solubility. The similar K_M and k_{cat} values obtained for all the Cbz-derivatized Gln-Xaa peptides indicate that the identity of the C-terminal amino acid does not significantly affect productive binding or turnover. Thus, these substrates result in a similar catalytic efficiency. Because amino acids having side chains with widely differing chemical and structural attributes were assayed and because we observed that the nature of these side chains does not affect substrate recognition or efficiency, we can conclude that substrate binding does not depend on the nature of the C-terminal amino acid. However, the nature of the N-terminal functional group, Boc or Cbz, has an important effect on substrate recognition by the enzyme. Namely, Boc-Gln-Gly does not serve as a donor substrate, whereas Folk and Cole (1966) observed that Cbz-Gln does act as a substrate of tissue type TGase, albeit a poor one, although Gln-Gly does not. Our kinetic data confirm that the presence of the Cbz group contributes importantly to

recognition of dipeptide substrates by tissue TGase, presumably by allowing better binding to the enzyme.

Table 1. Kinetic constants for guinea pig liver TGase with various N-terminally derivatized peptides

Substrate	k_{cat} (min^{-1})[a]	K_M (mM)	k_{cat}/K_M (min^{-1} mM^{-1})[a]
Cbz-Gln-Gly	96	3.2 ± 0.5	30
Cbz-Gln-Ala	51	1.9 ± 0.6	26
Cbz-Gln-Val	71	4.4 ± 0.4	16
Cbz-Gln-Leu	127	7 ± 1	18
Cbz-Gln-Phe	54	2.7 ± 0.8	20
Cbz-Gln-Ser	66	2.9 ± 0.6	23
Cbz-Gln-Gly-Gly	102	9.4 ± 0.5	11
Boc-Gln-Gly	—	> 40 [b]	—

[a] Values determined using the extinction coefficient measured for the Cbz-Gln-Gly anilide. The standard deviation on these values is ± 20 %.

[b] No reaction was observed for concentrations up to 40 mM, the limit of solubility for this dipeptide.

3.3.2 Modeling of substrate binding

To obtain more information regarding the mode of binding of acyl-donor substrates on tissue TGase, we attempted to model substrate binding using the coordinates of the crystal structure of tissue TGase from red sea bream that were obtained at 2.5 Å resolution. This structure, although lacking the activating calcium ion, was chosen over the crystal structure of human tissue TGase because the latter was cocrystallized with GDP and is therefore in the inactive state (Liu *et al.* 2002). The percentage of identity of the catalytic

domains of red sea bream TGase with guinea pig liver TGase is >55 %, and the percentage of homology is ~70 %. Furthermore, all of the active-site residues (W236, C272, H300, W329, H332, and Y515 of red sea bream tissue TGase; Figure 1) are conserved in all sequenced tissue TGases, justifying the use of the coordinates of fish transglutaminase for our modeling in the absence of coordinates for the guinea pig liver TGase. As a first step toward modeling the acyl-donor substrate binding on tissue TGase, we created all of the substrates given in Table 1 *in silico* with the BIOPOLYMER module of InsightII, and their structures were energy-minimized. All adopted a "Y"-shaped extended conformation, with the vertical line representing the Gln side chain and the two arms pointing upwards being the peptide backbone, with the Cbz or Boc group on one side and on the other, the C-terminal amino acid. These conformations were used as starting structures for automated and manual docking at the active-site of tissue TGase.

The crystal structure of red sea bream TGase presents a shallow, narrow cleft running diagonally on its surface, passing over the active-site (Figure 2). We identified this cleft as the possible binding site of the small peptide substrate's backbone and N-terminal functional group, as it permitted maximal Gln side chain insertion into the active-site, bringing the γ-carboxamide group into close proximity with the catalytic triad. Visual inspection suggests that it would be difficult for it to be located in any significantly different manner because of steric clashes.

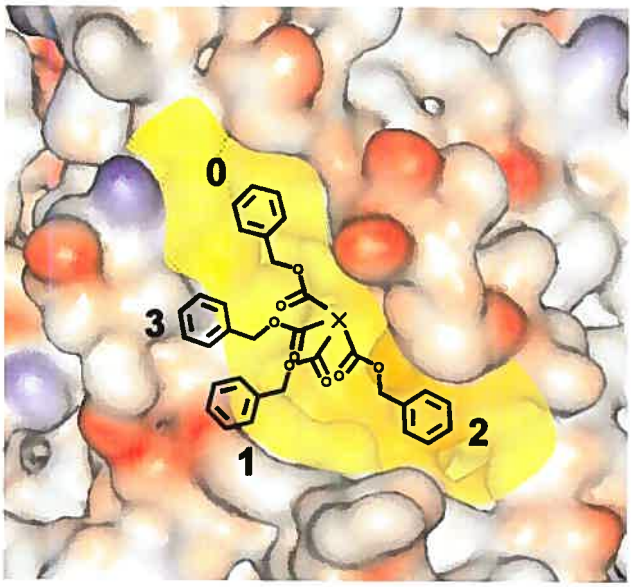


Figure 2. Proposed binding cleft for the peptide substrate N-terminal functional group at the surface of red sea bream TGase. The diagonal cleft running over the active-site cavity is in yellow; the Gln side chain of the Cbz-Gln-Gly substrate that enters the active-site cavity away from the viewer is represented by x. Negative charges are in red; positive are in blue. The four positions tested for manual docking of the N-terminal Cbz group are numbered. The Cbz group is drawn to scale.

To test this hypothetical binding mode, we performed automated docking experiments using AutoDock 3.0 software (Morris *et al.* 1998). Preliminary docking experiments were performed using the PDB file 1G0D. We observed that the Cbz-Gln-Gly and Boc-Gln-Gly substrates never entered the active-site during the 50 runs performed. Indeed, the active-site of the enzyme is visibly blocked by a Trp residue, W329 (Figure 1) which is conserved in all tissue TGases as well as in Factor XIII (Yee *et al.* 1994), epidermal (Ahvazi *et al.* 2002), and keratinocyte TGases. This apparently results from the structure having been resolved in the absence of substrate. Because automated docking is performed without movement of the protein atoms, it is unlikely to succeed in identifying a binding mode for the substrate Gln side chain inside the active-site. Thus, we manually

increased the accessibility of the active-site of the structure of red sea bream TGase by introducing torsions in the C_α–C_β (-83.13°) and C_β–C_γ (+48.52°) bonds of residue W329. These torsions opened the active-site cavity while minimizing steric clashes between W329 and the rest of the protein. With this starting structure, the AutoDock procedure resulted in several structures with the substrate Gln entering the active-site (Table 2). This indicates that W329 may prevent entry to the active-site, acting as a 'gate' that could regulate TGase activity. Of the 50 runs performed with Cbz-Gln-Gly, only four generated a structure where the substrate's Gln side chain had entered the active-site (Table 2). This low number of hits is expected because the substrate was allowed a high degree of conformational flexibility. Three of the four bound structures positioned the Cbz group at position 0; the last one positioned the Cbz at position 1 (Figure 2). With the Boc-Gln-Gly substrate, only six of the 50 runs generated a structure where the Gln side chain had entered the active-site. Of these six structures, two had the Boc at position 2, three at position 0, and one at position 1. It should be noted that these positions are not precisely defined but represent restricted zones within which positioning was observed. These AutoDock results allowed us to establish three potential binding sites at the surface of the enzyme for the N-terminal functional group. Position 0, which is in the diagonal cleft, appears to be favoured for binding. This suggests that positioning of the N-terminal functional group in the cleft is favourable to productive binding of the Gln side chain inside the active-site cavity. The low number of runs that resulted in a structure where Gln was bound in the active-site reflects the generally weak affinity of tissue TGase for these dipeptide substrates. This in turn is manifested in their millimolar K_M values. We also observed that after energy minimization of these 10 structures using InsightII, 75 % (3/4) of the Cbz-Gln-Gly structures formed an H-bond with conserved residue Y515 of the active-site, whereas only 16 % (1/6) of the Boc-Gln-Gly structures had formed this H-bond (Table 2).

Table 2. Automated docking results for the docking of Cbz-Gln-Gly and Boc-Gln-Gly on red sea bream TGase

Ligand	N[a]	Interaction Energy[b] (kcal/mol)	FG Position[c]	H-bond with Y515
Cbz-Gln-Gly	4	-56.6	1	√
		-21.9	0	×
		+23.3	0	√
		-8.4	0	√
Boc-Gln-Gly	6	-19.9	2	×
		+10.8	0	×
		+3.1	0	×
		-13.9	0	√
		-6.3	2	×
		-58.1	1	×

[a] Number of runs out of 50 that generated docking of the substrate with the Gln side chain entering the active-site.

[b] Total interaction energy between the enzyme and the ligand after energy minimization (see Methods).

[c] N-terminal functional group position according to Figure 2.

To further test the hypothesis that the binding site of the N-terminal functional group is in the cleft defined by positions 0 and 2 and to identify the preferred position for the N-terminal functional group in the cleft, we manually docked the acyl-donor substrate at four different positions on the enzyme. First, the reacting Gln side chain was positioned inside the active-site respecting the conditions described in the Materials and Methods section, while the N-terminal functional group was positioned in one of the four different positions on the surface of the enzyme (Figure 2). These positions place the N-terminal

functional group alternatively in the cleft (positions 0 and 2), in a small cavity perpendicular to the cleft (position 1), or directly on top of residue W236 (position 3). Steric clashes were more important in position 3, which was chosen as a negative control for the molecular modeling procedure. Then, steric clashes between the enzyme and the substrate were energy-minimized while the enzyme/substrate distance was constrained, as described in Materials and Methods. This caused the active-site to 'open up' as residue W329 was displaced, indicating that this residue is very mobile, because it readily moves away from the active-site after merely an energy minimization.

A 10-picosecond molecular dynamics simulation was then performed on the generated enzyme-substrate complexes to explore the immediate conformational space. One hundred conformers were thus generated. The conformer with the lowest energy was further energy-minimized, and the resulting structure was analyzed. This resulting structure with each substrate was always of lower energy than the structure generated before the molecular dynamics simulation. The backbone of the enzyme was maintained fixed during Trial 1 of manual docking (Table 3) to determine the most plausible binding site for the N-terminal functional group in order for the small Cbz-Gln-Gly substrate to adopt a conformation that best fit the enzyme's structure and not the opposite. This experiment was also performed two more times with varying conditions, including independent manual substrate positioning, a 25 Å or 40 Å layer of water covering the active-site, and with no constraint on the enzyme backbone, which generated similar results (Table 3). In all three independent trials with Cbz-Gln-Gly, the enzyme-substrate interaction energy was the lowest and was consistently negative when the Cbz was positioned in the upper left part of the cleft (position 0; Figure 2), indicating that it is the only one of the four positions tested that is energetically favourable and reproducible for binding the Cbz group for a small peptide substrate. Thus, it appears that the preferred position for docking the Cbz group on the surface of the enzyme is located in the upper left corner of the diagonal cleft (position 0). This position also allows better insertion of the reacting Gln side chain into the active-site based on the *anti*-conformation of the Gln side chain in the active-site compared to the

other three positions (data not shown). Manual docking experiments therefore gave us similar results to those obtained with automated docking, while providing additional insights into the preferred position of the N-terminal functional group. The remainder of the docking experiments was performed with manual docking of the substrates on tissue TGase.

Table 3. Manual docking results for Cbz-Gln-Gly on red sea bream TGase

Trial	FG Position[a]	Interaction Energy[b] (kcal/mol)
1[c]	0	-21.1
	1	+36.5
	2	+48.0
	3	+94.4
2[d]	0	-32.9
	1	+38.2
	2	-26.6
	3	-30.0
3[e]	0	-22.8
	1	+28.8
	2	+15.1
	3	+74.8

[a] N-terminal functional group position as defined in Figure 2.

[b] Total interaction energy between the enzyme and the ligand after energy minimization (see Methods).

[c] 25 Å layer of water covering the active-site, protein backbone was fixed.

[d] 25 Å layer of water covering the active-site, no fix on the protein backbone.

[e] 40 Å layer of water covering the active-site, no fix on the protein backbone.

Following the docking studies on the putative positioning of the Cbz group using Cbz-Gln-Gly, binding of the Cbz-Gln-Xaa substrates was modeled with the N-terminal functional group in position 0. The Boc-Gln-Gly substrate was modeled as a control to ensure that the modeling results are consistent with its lack of reactivity as a substrate (Table 1). For each test, the substrate was manually positioned in an independent manner. During the calculations, only the residues including atoms within 20 Å of the catalytic S_γ atom of C272, the water molecules, and the substrate were mobile. Constraints were applied as described in Materials and Methods to retain the substrate Gln proximal to the catalytic residues, thus increasing the likelihood of modeling productive enzyme/substrate complexes. After the energy minimization/molecular dynamics methodology was performed, the interaction energy between the enzyme and each substrate was analyzed. For the Cbz-Gln-Gly substrate, analyzed in 13 tests, we found that the interaction energy (ΔG) between ligand and protein was negative, thus favourable, in all but one case (Table 4). In the single case where the interaction energy was positive, the electrostatic component of the interaction energy between enzyme and substrate was very high (+158.9 kcal/mole). In this case, the carboxylate of Cbz-Gln-Gly was found to be in closer proximity than usual to a conserved, negatively charged residue of red sea bream TGase, Glu360, which could result in such an unfavourable electrostatic energy. Nonetheless, the average of the 13 data sets suggests that there is a favourable interaction between the enzyme and the Cbz-Gln-Gly substrate. The average van der Waals energy between the enzyme and the Cbz-Gln-Gly ligand was -53 ± 5 kcal/mole. While examining the interactions between the enzyme and Cbz-Gln-Gly, we found that each of the 13 structures generated showed an H-bond between the O_η proton of Y515 and the O_δ of Gln from Cbz-Gln-Gly. Also, 77 % of the structures (10/13) displayed an H-bond between N_ε of H332 and H_ε of Gln side chain. The proximity and orientation displayed in this H-bond pairing resembles the interaction that takes place during catalysis when the acid/base catalyst, H332, protonates the ammonia leaving group. The high prevalence of these H-bonds suggests that they are important in binding the glutaminyl residue at the active-site.

Table 4. Manual docking results for Cbz-Gln-Gly, Cbz-Gln-Xaa and Boc-Gln-Gly[a]

Ligands	H-bond with Y515 (%)	H-bond with H332 (%)	Van der Waals E (kcal/mol)	$E^b \geq 0$ kcal/mol (%)
Cbz-Gln-Gly	100 (13/13)	77 (10/13)	-53 ± 5	8 (1/13)
Cbz-Gln-Xaa[c]	100 (7/7)	86 (6/7)	-60 ± 5	0 (0/7)
Boc-Gln-Gly	69 (9/13)	62 (8/13)	-44 ± 4	38 (5/13)

[a] Ligands were positioned with the N-terminal functional group at position 0, as defined in the text.

[b] Total interaction energy between the enzyme and the ligand that is positive, thus unfavourable.

[c] Xaa = Ala, Val, Leu, Phe, Ser or Gly-Gly. Ala was modelled twice.

For the Cbz-Gln-Xaa substrates (Xaa = Ala, Val, Leu, Phe, Ser, or Gly-Gly), we obtained similar results. Interaction energies between the enzyme and the substrates were negative and thus favourable for all substrates, independently of the side chain of the C-terminal amino acid. Also, all of these molecules formed the same H-bond between the proton of O_η of Y515 and the O_δ of Gln, as observed for Cbz-Gln-Gly, whereas 86 % (6/7) of them formed an H-bond between N_ε of H332 and H_ε of the Gln side chain. These results indicate that the nature of the second amino acid does not significantly affect efficiency of binding to tissue TGase, which correlates well with the kinetic data from Table 1.

To evaluate the contribution of the N-terminal functional group to substrate binding, we compared the binding of Boc-Gln-Gly and Cbz-Gln-Gly to tissue TGase. The Boc-Gln-Gly peptide was manually docked similarly to the Cbz-Gln-Gly peptide with the N-terminal functional group at position 0 or position 2. The results at position 2 were similar to those for Cbz-Gln-Gly at position 2; that is, they were of considerably higher energy and were not pursued (data not shown). The binding of Boc-Gln-Gly at position 0

did not generate reproducible results as Cbz-Gln-Xaa did. Indeed, some results generated poor interaction energy between enzyme and substrate, whereas others gave as good an interaction as the Cbz peptides. Of the 13 trials, 38 % (5/13) generated a positive interaction energy, indicating unfavourable binding. Furthermore, the H-bond between the Boc-Gln-Gly substrate and Y515 was found in only 69 % (9/13) of the trials, whereas the H-bond with H332 was found in 62 % of the trials (8/13), which is significantly less than for the Cbz-derivatized peptides. In addition, the average van der Waals energy is -44 ± 4 kcal/mole, which is more than 10 kcal/mole higher than for the average of all Cbz-derivatized peptides (Cbz-Gln-Gly and Cbz-Gln-Xaa). Taken together, these results suggest that the presence of the Boc group does not allow for binding that is as constant and reproducible as for the Cbz-Gln-Xaa substrates. It thus appears that the Boc group is deleterious to proper binding, because it does not allow a constant productive insertion of Gln side chain into the active-site.

3.3.3 Modeling of the tetrahedral and acyl-enzyme intermediate

To minimize any potential bias arising from the manual docking undertaken during the modeling study and to investigate more thoroughly the importance of the H-bond network with Y515 and H332 identified for the binding of the Cbz-peptides, we created *in silico* the tetrahedral and acyl-enzyme intermediates for the reaction of tissue TGase with Cbz-Gln-Gly. These intermediates have a covalent bond between the C_δ atom of Gln from the donor substrate and S_γ of the catalytic Cys residue, thus eliminating the requirement for manual docking of the substrate. The acyl-enzyme and tetrahedral intermediates were constructed respecting the $\chi1$–$\chi2$ and $\chi2$–$\chi3$ plots dihedral angles (Janin *et al.* 1978) for the substrate Gln and C272. Two starting structures for the acyl-enzyme intermediate were constructed, one with the Cbz group at position 0 and another at position 2. These dihedral angles still respected the $\chi1$–$\chi2$ and $\chi2$–$\chi3$ plots at the end of the simulation, indicating that the generated structure was coherent with regards to the conformation of the two residues involved in the covalent bond between enzyme and substrate (data not shown). After the

simulation, a lower energy for the modeled system was obtained for the acyl-enzyme intermediate with the Cbz group at position 0 compared to position 2 (-4513 kcal/mole versus -4461 kcal/mole). This is consistent with the hypothesis that position 0 represents the actual Cbz binding site because it is energetically more favorable.

While analyzing the covalent structures generated, we observed that the acyl portion of the molecule adopted a conformation very similar to the model of TGase with the noncovalently docked substrate. Indeed, the RMSD of the Gln side chains of Cbz-Gln-Gly in the Michaelis complex and the acyl portion of the acyl-enzyme and tetrahedral intermediates are very low (Table 5), indicating near-identical positioning of the Gln side chain inside the active-site for these three modeled structures (Figure 3A–C). Also, the dihedral angles $\chi 1$ and $\chi 2$ are nearly identical among the three structures, which demonstrates that the conformation of the Gln side chain in the Michaelis complex is the same as in the tetrahedral and acyl-enzyme intermediates. This suggests that the insertion of the Gln side chain of the docked substrate is coherent and thus there was no bias in positioning during manual docking. The H-bond between the reactive Gln γ-carboxamide group and the enzyme Y515 residue was present in the acyl-enzyme intermediate. The H-bond between N_ε of H332 and H_ε of Gln side chain was not present because the Gln side chain in the acylenzyme intermediate no longer carries the NH_2 group.

Table 5. Comparison of the conformation of the Gln side chain of the acyl-donor substrate Cbz-Gln-Gly in the modelled Michaelis complex, acyl-enzyme and tetrahedral intermediates

	RMSD of Gln side chain (Å)			Dihedral angles of Gln	
Structure	Acyl-enzyme intermediate	Tetrahedral intermediate	Michaelis complex	$\chi 1$ (°)	$\chi 2$ (°)
Acyl-enzyme intermed.	-----	0.12	0.10	-175	+172
Tetrahedral intermed.	0.12	-----	0.02	-178	-169
Michaelis complex	0.10	0.02	-----	-176	-171

There are two H-bonds in the tetrahedral intermediate structure (Figure 3B) that are similar to the H-bonds between Cbz-Gln-Gly and the enzyme in the Michaelis complex. The first H-bond occurs between the oxyanion generated by nucleophilic attack and the hydroxyl function of Y515. The second occurs between N_ε of Gln and the acidic proton of protonated H332. This is consistent with the known mechanism of the enzyme, where the leaving group NH_2 is protonated by the imidazolium group of H332. The H-bond formed between the oxyanion and Y515 could significantly stabilize the tetrahedral intermediate. These results, along with the docking study of Cbz- or Boc-derivatized peptides, support the importance of these H-bonds.

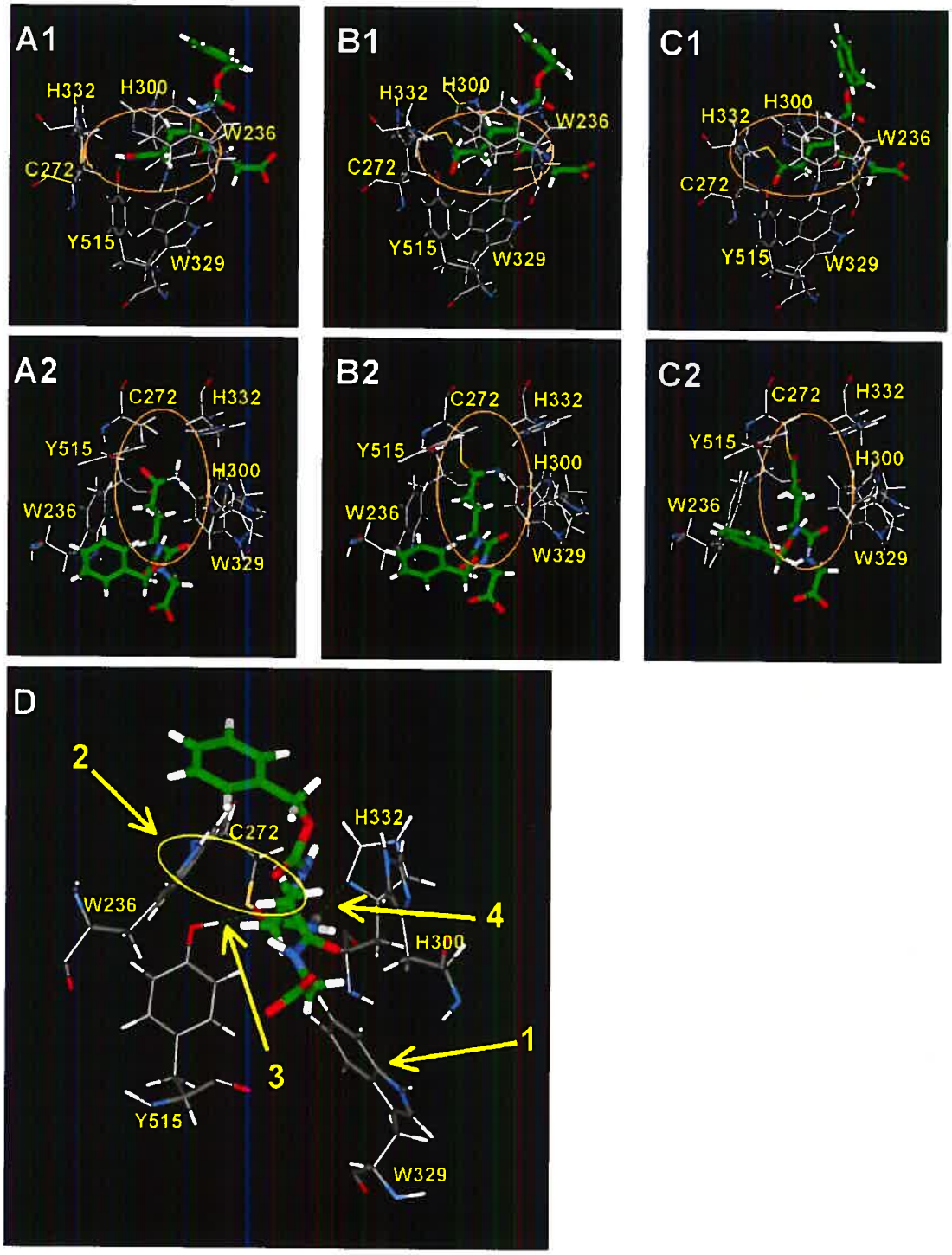


Figure 3. Model of TGase with the acyl-donor Cbz-Gln-Gly. (A) Michaelis complex, (B) tetrahedral intermediate, and (C) acyl-enzyme intermediate. A1–C1, side view. A2–C2, top

view. In C2, the NH_3 molecule has been released. (D) Interactions proposed to be important in the modeled Michaelis complex of TGase with Cbz-Gln-Gly: (1) Displacement of W329 toward the bottom to allow entrance of substrate Gln side chain into the active-site; (2) Hydrophobic interaction of the indole group of W236 with the methylene groups of the Gln side chain; (3) H-bond of the Gln γ-carboxamide with Y515; (4) H-bond of the Gln γ-carboxamide with H332.

3.3.4 Structural analysis of Cbz-peptide binding

The 13 structures generated by molecular dynamics following manual docking of Cbz-Gln-Gly peptides and the seven structures generated with the Cbz-Gln-Xaa substrates were analyzed as a group. The Gln side chain enters the active-site almost perpendicularly with respect to the center of mass of the enzyme. Figure 3D illustrates one of the trial structures, chosen because it was typical and representative. The conformation of the substrate Gln side chain is the low-energy *anti*-conformation. The $\chi 1$ and $\chi 2$ dihedral angles of the substrate Gln are near 180°, which gives a *trans*-conformation to the side chain that is common for amino acids. Three aromatic residues, W236, W329, and H300 that appear to form a tunnel leading to the catalytic residues, surround the side chain. As previously discussed, residue W329 is displaced by the insertion of the substrate Gln side chain into the active-site so as to open the active-site that it blocks in the unbound state. This displacement places the W329 side chain in a position very similar to that achieved manually as part of the automated docking experiments (C_α-C_β = -95.53° and C_β-C_γ = +82.4°). The roof of the tunnel is formed by a loop of residues P356–G369 (Figure 1). The two methylene groups of the Gln side chain of the substrate are in contact with the indole group of residue W236. This maximizes the hydrophobic interaction between the indole group of W236 and the methylene groups of the substrate Gln side chain during the course of the simulation—as evidenced by the concerted movement of these two groups during the dynamics simulation. The substrate γ-carboxamide, as mentioned earlier,

adopts a conformation allowing it to form H-bonds with residues Y515 and H332. In this conformation, the nucleophilic attack by the thiolate can only be achieved from the top or the bottom of the substrate γ-carboxamide, in keeping with an addition-displacement mechanism. If the attack occurs from the top, the oxyanion generated could form a stabilizing H-bond with the hydroxyl group of Y515. Such a bond is observed in the model of the tetrahedral intermediate.

3.4 Discussion

No existing crystal structures of tissue TGase (Noguchi *et al.* 2001; Liu *et al.* 2002) have been resolved with ligand bound at the active-site, limiting our understanding of the binding of acyl-donor substrates. The only structural elements that have been shown to be important for productive binding of small peptide acyl-donor substrates are (1) the length of the side chain (Asn versus Gln), (2) the stereochemistry of the Gln side chain (D-Gln versus L-Gln), (3) the length of the peptide substrate (Clarke *et al.* 1959; Folk 1983), and (4) to a lesser extent, the position of the reactive Gln in the peptide/protein sequence (Ohtsuka *et al.* 2000). Thus, it is known that for small acyl-donor peptide substrates, the reactive Gln residue must be at least the second residue from the C terminus and the third from the N terminus (Folk and Cole 1965), where Cbz can replace the first two amino acids. Molecules such as Gln and Gln-Gly do not act as substrates, whereas Cbz-Gln-Gly is the most widely used substrate.

To gain insight into the binding of small peptide acyl-donor substrates on tissue TGase, we used a combined kinetic/molecular modeling approach. Kinetic tests were performed on guinea pig liver transglutaminase, because it is the most extensively characterized tissue TGase. Molecular modeling was performed on the crystal structure of red sea bream tissue TGase. The catalytic core domain of this enzyme has a sequence homology of ~70 % to that of guinea pig liver, and all but four residues composing the cleft (Figure 2) are conserved between the two enzymes, and thus the two structures can be

considered very similar. All of the active-site residues (W236, C272, H300, W329, H332, and Y515 of red sea bream tissue TGase) are conserved in all tissue TGases, including the human enzyme, of which the catalytic domain has 58 % identity and 70 % homology with the red sea bream enzyme. This justifies the use of the coordinates of red sea bream transglutaminase for our modeling in the absence of coordinates for the guinea pig liver TGase, as well as the pertinence of these studies to future drug design. Furthermore, Cbz-Gln-Gly is an efficient acyl-donor substrate of both red sea bream TGase and guinea pig liver TGase (Ohtsuka *et al.* 2000) and is therefore appropriate for docking studies on this structure.

Molecular modeling suggests that there is a hydrophobic interaction between the indole group of the side chain of W236 and the methylene groups of the Gln side chain of the substrate. Molecular dynamics simulations also confirm the concerted movement of these two groups. Residue W236 (W241 in guinea pig liver TGase and human tissue TGase) is conserved in all eukaryotic TGases that have been sequenced. In site-directed mutagenesis studies of human tissue TGase, replacement of W241 with Ala, Gln, His, Phe, or Tyr resulted in a substantial loss of activity, demonstrating the importance of this Trp to catalysis (Murthy *et al.* 2002). Significantly, the mutants W241H, W241F, and W241Y, possessing aromatic side chains, showed low but readily measurable activities of approximately one-tenth that of wild-type and around 20-fold higher than that of the non-aromatic mutants W241A and W241Q (Murthy *et al.* 2002). In consideration of these results and the contacts revealed by our modeling studies of red sea bream TGase, we propose that the acyl-donor substrate Gln residue is stabilized in the active-site through hydrophobic interactions of the methylene groups of the substrate side chain with the aromatic indole group of residue W236. This could partly account for the fact that peptide- or protein-bound Asn is not a substrate of TGase because its shorter and less hydrophobic side chain, lacking one methylene unit compared to Gln, would not allow these stabilizing interactions.

Molecular modeling suggests that residue Y515 is crucial in acyl-donor substrate binding and reactivity. The presence of a ubiquitous H-bond between the hydroxyl group of Y515 and the carbonyl oxygen from Gln side chain for all of the efficient substrates (Cbz-peptides) indicates that Y515 participates in stabilizing the substrate in the active-site. This tyrosine residue is strictly conserved in all tissue TGases. It belongs to the barrel 1 domain but in the folded structure is located within the active-site and in close proximity to the catalytic C272 in the core domain. Taken together, these data suggest that Y515 plays an important role in this enzyme. It has been proposed to form a regulatory H-bond with the catalytic sulphur atom, preventing nucleophilic attack by the sulphur atom on the substrate carboxamide group (Yee *et al.* 1994; Noguchi *et al.* 2001). However, our findings suggest that it could orient the substrate γ-carboxamide for nucleophilic attack by the thiolate and that the H-bond formation could activate the carbonyl towards this nucleophilic attack by rendering it more electrophilic. Our studies also reveal the presence of this H-bond in the structures of the acyl-enzyme and tetrahedral intermediates, further suggesting its involvement in stabilizing the oxyanion generated during catalysis. We are currently undertaking mutagenesis of guinea pig liver TGase at this position to verify our hypothesis.

The H-bond network identified in our modeling study, involving residues Y515 and H332 of red sea bream tissue TGase and the γ-carboxamide group of the donor substrate, could be critical for transglutaminase activity. This network is highly prevalent in the structures generated by modeling with efficient substrates (Cbz-peptides) and is present at a significantly lower percentage in the structures generated with the unreactive dipeptide Boc-Gln-Gly. This H-bond network could catalyze acylation by orienting the γ-carboxamide group relative to the two catalytic residues C272 and H332, activating the γ-carbonyl group to nucleophilic attack by C272 by rendering it more electrophilic and stabilizing the oxyanionic transition state. Molecular modeling showed that the H-bond between the O_η proton of Y515 and the O_δ of Gln from Cbz-Gln-Gly was present 100 % of the time with all efficient substrates as well as in the acyl-enzyme and tetrahedral

intermediates. However, this H-bond was not always present when a poor substrate, Boc-Gln-Gly, was modeled.

Residue W329 blocks the active-site in the crystal structure of red sea bream TGase and human TGase (W332 in human tissue TGase), preventing substrate entry. Indeed, automated docking experiments with AutoDock 3.0 using the crystal structure of the red sea bream enzyme whose active-site is blocked by residue W329 did not generate structures with substrate in the active-site. However, we have shown by molecular dynamics simulation over 10 ps that this residue readily moves to allow insertion of the side chain into the active-site. This tryptophan residue is also strictly conserved in all of the tissue TGases and is located in a loop comprising residues S321 to W329. It could therefore act as a 'gate' that could close the active-site and exclude water. W329 could then move to open the active-site upon substrate binding, allowing substrates to enter. The opening of the active-site by displacement of this residue may be mediated by allosteric calcium binding, which is essential for TGase activity. This hypothesis is indirectly supported by results from crystallographic studies of TGase 3, showing that calcium binding induces conformational changes that expose the corresponding W327 in a way that should increase its dynamic mobility and interactions with incoming substrates (Ahvazi *et al.* 2002).

The difference in recognition reflected by the varied K_M values for Gln-Gly dipeptides having different N-terminal functional groups, Boc or Cbz, could be attributed to various factors. First, the van der Waals energy of interaction between the enzyme and the N-terminally derivatized Gln-Gly dipeptide substrate varies by almost 10 kcal/mole, depending on the N-terminal functional group. This difference in interaction energy correlates well with the poorer binding of the Boc-derivatized dipeptides. Furthermore, modeling of the binding of Boc-derivatized dipeptides gave unfavorable, positive interaction energies far more frequently than Cbz-derivatized dipeptides. Finally, the aforementioned H-bond network is frequently absent in the structures generated with the Boc-dipeptides. These differences in binding could be the result of differences in the shape

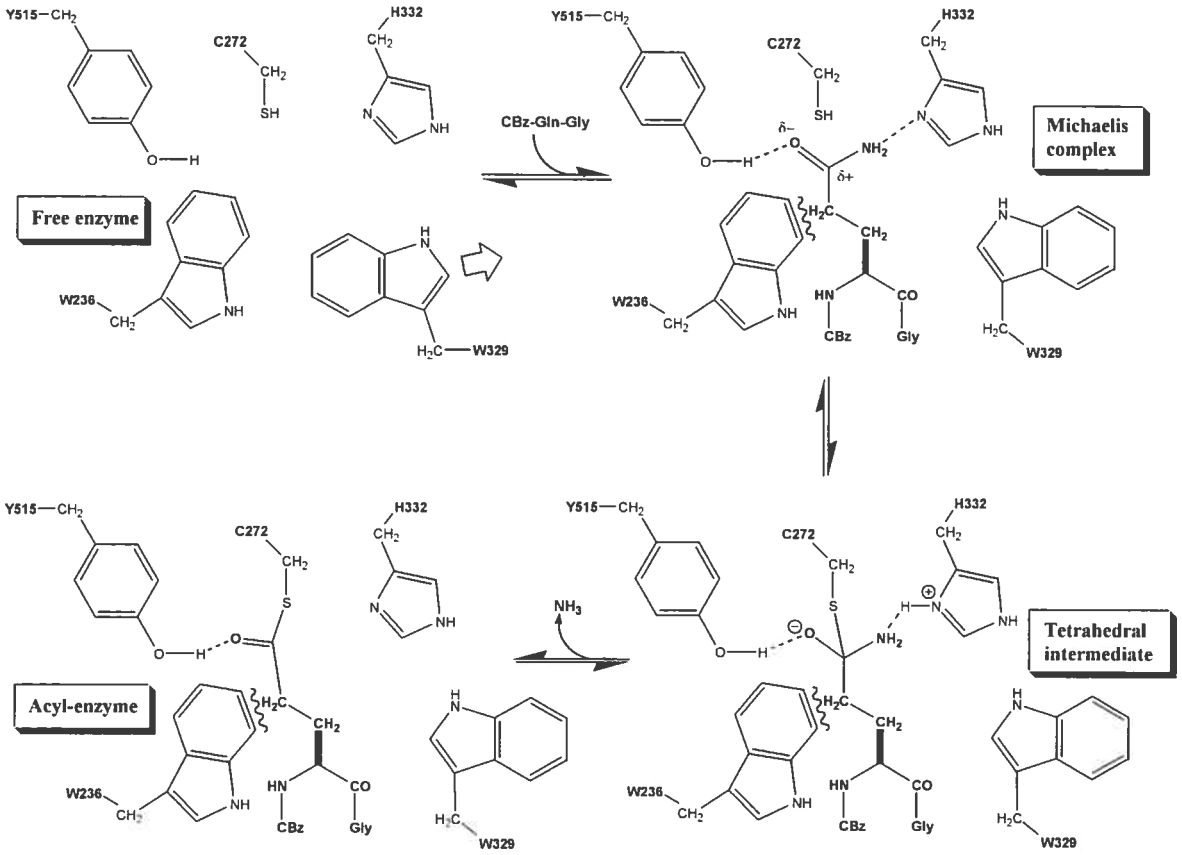
and size between the Cbz and Boc groups, because the Cbz is planar whereas the Boc is globular. The Boc group is also shorter in that the bulky *tert*-butyl group is attached directly to the carbamate oxygen, whereas the bulky phenyl group of the Cbz group is attached to an additional methylene unit, increasing its length and conformational flexibility. It is also possible that good recognition of Cbz-Gln-Gly by tissue TGase is conferred by the aromaticity of the Cbz group. We are currently undertaking kinetic studies of various substrates containing differing N-terminal functional groups to explore the importance of the nature of this group to binding affinity.

Summarized in Table 6 and Scheme 2 are the interactions that we propose for the acylation of tissue TGase by small peptide substrates. First, the active-site is opened by displacement of residue W329. This displacement could occur after calcium binding and/or upon binding of acyl-donor substrate to the enzyme. The Gln side chain then enters the active-site. Its insertion is favoured and stabilized by hydrophobic interactions between the methylene groups of the Gln side chain and the indole group of residue W236. This residue could also be partly responsible for the specificity towards Gln rather than Asn. The γ-carboxamide group of the substrate Gln then forms the H-bond network with residues Y515 and H332. This network orients the γ-carboxamide group for nucleophilic attack by the catalytic thiol and increases electrophilicity of the γ-carboxamide, making it more reactive. After nucleophilic attack, the resulting oxyanion is stabilized by the H-bond formed with the hydroxyl group of Y515, while the NH_2 group of the γ-carboxamide forms an H-bond with the acidic proton of the H332 imidazolium. Then, ammonia is released on formation of the acyl-enzyme intermediate.

Table 6. Proposed interactions between acyl-donor substrates and tissue TGase

INTERACTION	POSSIBLE ROLE
Steric obstruction by W329	• 'Gate' to the active-site
Hydrophobic interaction with W236	• Stabilization of Gln side chain insertion into active-site • Specificity for Gln side chain vs. Asn
H-bond with H332	• Orientation of substrate γ-$CONH_2$ group for nucleophilic attack
H-bond with Y515	• Orientation of substrate γ-$CONH_2$ group for nucleophilic attack • Activation of the substrate γ-carbonyl toward nucleophilic attack • Stabilization of the tetrahedral intermediate

Using a combined kinetic/molecular modeling approach, we generated a model for the binding of small acyl-donor peptide substrates on tissue TGase, and identified possible structural elements important for productive binding. Our findings, while correlating well with the known mechanism and specificity of tissue TGases, increase our understanding of the binding of acyl-donor substrates on tissue TGase. While raising new hypotheses regarding the importance of the N-terminal functional group of small peptide substrates, our model should be helpful in the future development of inhibitors for this class of enzymes, implicated in various diseases.



Scheme 2. Acylation step of the TGase mechanism.

3.5 Materials and methods

3.5.1 Synthesis of the N-terminally derivatized peptide substrates

Peptide substrates were synthesized according to a published procedure based on the activation of Cbz- or Boc-protected amino acids as their corresponding *p*-nitrophenyl esters, followed by reaction with unprotected amino acids (Gagnon *et al.* 2002).

3.5.2 Kinetic measurements

Kinetic parameters were measured using a kinetic protocol previously developed in our laboratory (de Macédo *et al.* 2000). According to this method, *N*,*N*-dimethyl-1,4-phenylene diamine (DMPDA) is used as an acceptor substrate for the TGase-mediated transamidation from Gln-containing donor substrates. For Cbz-Gln-Gly, the anilide resulting from the transamidation reaction with DMPDA has been shown to have an extinction coefficient of 8940 $M^{-1}cm^{-1}$ at 278 nm. Although the specific extinction coefficients of each of the corresponding anilides formed from the other substrates used in this study were not determined, they were assumed to not differ significantly. This was corroborated by the observation that similar concentrations of different anilide products gave similar absorbance values, and the k_{cat} values shown in Table 1 were calculated based on this assumption. A K_M value was measured for each substrate using this method at 37 °C and pH 7.0, over substrate concentrations that ranged from 0.2- to fivefold K_M. By way of comparison, a value of 3.2 mM was thus determined for the K_M of Cbz-Gln-Gly, in accordance with that determined in the literature by various methods (Folk and Chung 1985; Day and Keillor 1999; de Macédo *et al.* 2000).

3.5.3 Computational methods

Computations were performed using the InsightII package (version 2000, Accelrys). The BIOPOLYMER module was used to build or modify molecular structures, and all energy minimizations and molecular dynamics calculations were performed with the DISCOVER module using the consistent valence force field (CVFF). Interaction energies between the enzyme and the ligands were calculated with the DOCKING module of InsightII. Automated docking was performed with AutoDock 3.0 software (Scripps).

3.5.4 Preparation of the protein structure

The starting coordinates were taken from the PDB file 1G0D of red sea bream tissue transglutaminase (Noguchi *et al.* 2001). The crystallographic water molecules as well as the sulfate ion were removed. Hydrogen atoms were added using the BIOPOLYMER module at the normal ionization state of the amino acids at pH 7.0. Two dihedral angle torsions involving hydrogen atoms were introduced in the protein structure in order to render it more consistent with the catalytic process. The first involves the hydrogen atom of the hydroxyl group of Y515, which was turned towards the sulphur atom of the catalytic C272. This torsion was introduced according to the suggested existence of a regulatory H-bond between the catalytic Cys and Y515 of red sea bream tissue TGase (Noguchi *et al.* 2001). The second involves the hydrogen atom of the thiol group of C272, which was turned towards N_δ of H332. This torsion was introduced according to the known catalytic mechanism of the enzyme in which the general acid-base catalyst, H332, deprotonates the catalytic Cys residue prior to or during catalysis (Case and Stein 2003). Energy minimization of the added hydrogen atoms was then performed to remove bad contacts in the protein structure. One thousand steps of steepest descents energy minimization were performed, followed by a conjugate gradients energy minimization until convergence of 0.01 kcal/mole/Å. Both energy minimizations were performed while keeping the heavy atoms of the protein fixed and using a dielectric constant of 1. This generated the structure g0d300.

3.5.5 Computational construction of substrate molecules

Substrates were built using fragments from the BIOPOLYMER module fragment library. The resulting structures were energy-minimized using 1000 steps of steepest descents energy minimization followed by a conjugate gradients energy minimization until convergence of 0.01 kcal/mole/Å with a distance-dependent dielectric constant of 80 to mimic implicit water. These resulting molecules served as starting structures for the docking experiments.

3.5.6 Automated docking of substrates on TGase

AutoGrid was used to generate geometry-centered grid maps with 0.375 Å spacing based on a region that contained all residues including atoms within a 15 Å radius centered about the catalytic sulphur atom in the structure g0d300. A distance-dependent dielectric constant of 4.0 was used to mimic the interior of the protein. The genetic algorithm implemented in AutoDock 3.0 (Morris *et al.* 1998) was applied with a starting population size of 50. Fifty runs were performed for each ligand (Cbz-Gln-Gly and Boc-Gln-Gly) with a maximum number of energy evaluations of 250,000 and a maximum number of generations of 27,000. Auto-Tors was used to define the available torsions for the ligands. However, we restricted the torsions of the Gln side chain (C_α–C_β, C_β–C_γ and C_γ–C_δ) to define an extended conformation that maximizes its length. The torsions of the methyl groups of the Boc functional group were also deleted, as they do not affect the structure of the molecule. There remained seven possible torsions for Cbz-Gln-Gly and six for Boc-Gln-Gly. After automated docking, the resulting enzyme/substrate structures were energy-minimized using the DISCOVER module of InsightII. A simulation zone containing all residues including atoms within 20 Å of the sulphur atom of the catalytic cysteine residue was used, with the remainder of the protein being fixed. An assembly was created between the enzyme and the substrate, and a layer of 25 Å of explicit water was added around the substrate. This layer of water completely covered the active-site region. A nonbond cutoff of 11 Å was used to speed up calculations, and a distance restraint was applied to maintain the substrate's reactive group (γ-carboxamide group of Gln) at a maximal distance of 3 Å from the catalytic S_γ of C272 and N_δ of H332. A dielectric constant of 1 was applied. A multistage energy minimization was performed on the enzyme/substrate complex to further refine the complex. One thousand iterations of steepest descents energy minimization was performed, followed by a conjugate gradients energy minimization until convergence of 0.001 kcal/mole/Å.

3.5.7 Manual docking of substrates on TGase

Manual docking of the minimized substrates on TGase was performed as follows. First, the reactive Gln side chain of the rigid substrates was inserted into the well-known active-site cavity (Yee *et al.* 1994) perpendicularly to the center of mass of the protein so as to maximize the depth of insertion while maintaining all protein atoms fixed. The side chain was placed in a sterically unhindered position that is consistent with Gln side chain insertion into the active-site cavity. Second, the distance between the reactive atoms of the enzyme and the substrates (S_γ of C272 with C_δ of the Gln side chain and N_ε of Gln side chain with N_δ of H332) was kept at values lower than 3 Å. Third, the peptide backbone and the N-terminal functional group of the substrates were placed in a sterically unhindered position on the surface of the enzyme, with the functional group in one of the four tested positions (Figure 2). Finally, an assembly was created between the enzyme and the substrate, and a layer of either 25 Å or 40 Å of explicit water was added around the substrate. This layer of water completely covered the active-site region.

3.5.8 Calculations on the enzyme/substrate complex

For all reaction steps, a simulation zone containing all residues including atoms within 20 Å of the sulphur atom of the catalytic C272 residue was used, with the remainder of the protein being fixed. A nonbond cutoff of 11 Å was used to speed up calculations, and a distance restraint was applied to maintain the substrate's reactive group (γ-carboxamide group of Gln) close to the catalytic residues at a maximal distance of 3 Å from S_γ of C272 and N_δ of H332. For the molecular dynamics experiments, an additional restraint of 100 kcal/mole on the oxygen atoms of the water molecules was added to prevent them from "boiling off." A dielectric constant of 1 was applied.

A multistage energy minimization was performed on the enzyme/substrate complex to remove bad contacts between the substrate and the enzyme. First, 1000 iterations of

steepest descents energy minimization was performed, followed by a conjugate gradients energy minimization until convergence of 0.001 kcal/mole/Å. This first energy minimization was followed by a molecular dynamics simulation in which the molecular system was allowed to equilibrate at 300 K for 1 ps, followed by the actual simulation to explore conformational space for 10 ps while maintaining the same temperature. Snapshots were taken every 0.1 ps, generating 100 different conformers. The conformer with the lowest energy was energy-minimized with 100 steps of steepest descents followed by a conjugate gradients energy minimization to convergence of 0.001 kcal/mole/Å. At this point, the structures were analyzed.

3.5.9 Building of the acyl-enzyme intermediate and the tetrahedral intermediate

As a starting point for the energy calculations, the acyl-enzyme and tetrahedral intermediates were built using the BIOPOLYMER module. While building these molecules, we introduced torsions at the $\chi 1$, $\chi 2$, and $\chi 3$ dihedral angles of the catalytic C272 and the reactive Gln of the substrate. These angles ($\chi 1 = -30°$ and $\chi 2 = -60°$ for Cys, $\chi 1 = -160°$, $\chi 2 = 180°$, and $\chi 3 = 0°$ for Gln) were selected from published $\chi 1$–$\chi 2$ and $\chi 2$–$\chi 3$ plots for amino acid side chains (Janin *et al.* 1978). The angles allowed the positioning of the N-terminal functional group in position 0 as defined in Figure 2, whereas the angles $\chi 1 = -45°$ and $\chi 2 = -45°$ for Cys, $\chi 1 = 180°$, $\chi 2 = 180°$, and $\chi 3 = 180°$ for Gln allowed the positioning of the N-terminal functional group in position 2. For the tetrahedral intermediate, the catalytic His residue was protonated, a formal charge of -1 was introduced at the oxyanion, and the partial charges were calculated using the CVFF force field. These molecules were then minimized and subjected to a molecular dynamics simulation in the same manner as the enzyme/substrate intermediates.

3.6 Acknowledgments

We thank Dr. Andreea Schmitzer for assistance in molecular modeling. R.A.C. and P.G. were recipients of Fonds Québécois de recherche sur la nature et les technologies (FQRNT) fellowships. This research was funded by FQRNT Grant 80357 (jointly held by J.N.P. and J.W.K.).

3.7 References

Abe, S., Yamashita, K., Kohno, H., and Ohkubo, Y. 2000. Involvement of transglutaminase in the receptor-mediated endocytosis of mouse peritoneal macrophages. *Biol. Pharm. Bull.* **23:** 1511–1513.

Aeschlimann, D. and Paulsson, M. 1994. Transglutaminases: Protein cross-linking enzymes in tissues and body fluids. *Thromb. Haemost.* **71:** 402–415.

Ahvazi, B., Kim, H.C., Kee, S.H., Nemes, Z., and Steinert, P.M. 2002. Three-dimensional structure of the human transglutaminase 3 enzyme: Binding of calcium ions changes structure for activation. *EMBO J.* **21:** 2055–2067.

Case, A. and Stein, R.L. 2003. Kinetic analysis of the action of tissue transglutaminase on peptide and protein substrates. *Biochemistry* **42:** 9466–9481.

Chen, J.S. and Mehta, K. 1999. Tissue transglutaminase: An enzyme with a split personality. *Int. J. Biochem. Cell Biol.* **31:** 817–836.

Clarke, D.D., Mycek, M.J., Neidle, A., and Waelsh, H. 1959. The incorporation of amines into proteins. *Arch. Biochem. Biophys.* **79:** 338–354.

Day, N. and Keillor, J.W. 1999. A continuous spectrophotometric linked enzyme assay for transglutaminase activity. *Anal. Biochem.* **274:** 141–144.

de Macédo, P., Marrano, C., and Keillor, J.W. 2000. A direct continuous spectrophotometric assay for transglutaminase activity. *Anal. Biochem.* **285:** 16–20.

Fesus, L. and Piacentini, M. 2002. Transglutaminase 2: An enigmatic enzyme with diverse functions. *Trends Biochem. Sci.* **27:** 534–539.

Folk, J.E. 1969. Mechanism of action of guinea pig liver transglutaminase. VI. Order of substrate addition. *J. Biol. Chem.* **244:** 3707–3713.

———. 1983. Mechanism and basis for specificity of transglutaminase-catalyzed epsilon-(γ-glutamyl) lysine bond formation. *Adv. Enzymol. Relat. Areas Mol. Biol.* **54:** 1–56.

Folk, J.E. and Chung, S.I. 1973. Molecular and catalytic properties of transglutaminases. *Adv. Enzymol. Relat. Areas Mol. Biol.* **38:** 109–191.

———. 1985. Transglutaminases. *Methods Enzymol.* **113:** 358–375.

Folk, J.E. and Cole, P.W. 1965. Structural requirements of specific substrates for guinea pig liver transglutaminase. *J. Biol. Chem.* **240:** 2951–2960.

———. 1966. Transglutaminase: Mechanistic features of the active-site as determined by kinetic and inhibitor studies. *Biochim. Biophys. Acta* **122:** 244–264.

Gagnon, P., Huang, X., Therrien, E., and Keillor, J.W. 2002. Peptide coupling of unprotected amino acids through *in situ p*-nitrophenyl ester formation. *Tetrahedron Lett.* **43:** 7717–7719.

Huo, J., Metz, S.A., and Li, G. 2003. Role of tissue transglutaminase in GTP depletion-induced apoptosis of insulin-secreting (HIT-T15) cells. *Biochem. Pharmacol.* **66:** 213–223.

Janin, J., Wodak, S., Levitt, M., and Maigret, B. 1978. Conformation of amino acid side-chains in proteins. *J. Mol. Biol.* **125:** 357–386.

Kim, S.Y., Grant, P., Lee, J.H., Pant, H.C., and Steinert, P.M. 1999. Differential expression of multiple transglutaminases in human brain. Increased expression and cross-linking by transglutaminases 1 and 2 in Alzheimer's disease. *J. Biol. Chem.* **274:** 30715–30721.

Liu, S., Cerione, R.A., and Clardy, J. 2002. Structural basis for the guanine nucleotide-binding activity of tissue transglutaminase and its regulation of transamidation activity. *Proc. Natl. Acad. Sci.* **99:** 2743–2747.

Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., and Olson, A.J. 1998. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **19:** 1639–1662.

Murthy, S.N.P., Iismaa, S., Begg, G., Freymann, D.M., Graham, R.M., and Lorand, L. 2002. Conserved tryptophan in the core domain of transglutaminase is essential for catalytic activity. *Proc. Natl. Acad. Sci.* **99:** 2738–2742.

Noguchi, K., Ishikawa, K., Yokoyama, K., Ohtsuka, T., Nio, N., and Suzuki, E. 2001. Crystal structure of red sea bream transglutaminase. *J. Biol. Chem.* **276:** 12055–12059.

Ohtsuka, T., Ota, M., Nio, N., and Motoki, M. 2000. Comparison of substrate specificities of transglutaminases using synthetic peptides as acyl donors. *Biosci. Biotechnol. Biochem.* **64:** 2608–2613.

Priglinger, S.G., May, C.A., Neubauer, A.S., Alge, C.S., Schönfeld, C.L., Kampik, A., and Welge-Lussen, U. 2003. Tissue transglutaminase as a modifying enzyme of the extracellular matrix in PVR membranes. *Invest. Ophthalmol. Vis. Sci.* **44:** 355–364.

Shan, L., Molberg, Ø., Parrot, I., Hausch, F., Filiz, F., Gray, G.M., Sollid, L.M., and Khosla, C. 2002. Structural basis for gluten intolerance in celiac sprue. *Science* **297:** 2275–2279.

Shridas, P., Sharma, Y., and Balasubramanian, D. 2001. Transglutaminase-mediated cross-linking of α-crystallin: Structural and functional consequences. *FEBS Lett.* **499:** 245–250.

Singh, U.S., Pan, J., Kao, Y.L., Joshi, S., Young, K.L., and Baker, K.M. 2003. Tissue transglutaminase mediates activation of RhoA and MAP kinase pathways during retinoic acid-induced neuronal differentiation of SH-SY5Y cells. *J. Biol. Chem.* **278:** 391–399.

Wada, F., Nakamura, A., Masutani, T., Ikura, K., Maki, M., and Hitomi, K. 2002. Identification of mammalian-type transglutaminase in *Physarum polycephalum*.

Evidence from the cDNA sequence and involvement of GTP in the regulation of transamidating activity. *Eur. J. Biochem.* **269:** 3451–3460.

Yee, V.C., Pedersen, L.C., Le Trong, I., Bishop, P.D., Stenkamp, R.E., and Teller, D.C. 1994. Three-dimensional structure of a transglutaminase: Human blood coagulation factor XIII. *Proc. Natl. Acad. Sci.* **91:** 7296–7300.

CHAPITRE 4

Développement d'une méthode d'expression et de purification de la TGase de foie de cobaye chez *Escherichia coli*

4.0 Préface

Afin de pouvoir développer les méthodes expérimentales nécessaires à ce projet, il nous fallait une source de TGase de foie de cobaye recombinante. Pour ceci, nous avons conçu notre propre protocole d'expression et de purification de l'enzyme chez *Escherichia coli*. Lors du développement de ce protocole, nous avons dû nous attaquer au problème de faible solubilité intrinsèque de la TGase de foie de cobaye lorsqu'elle est exprimée chez *E. coli*. Ce protocole nous a permis d'obtenir rapidement, efficacement et avec un bon rendement la TGase de foie de cobaye soluble ayant l'activité spécifique la plus élevée rapportée dans la littérature. De plus, le fait de l'exprimer de manière recombinante offre la possibilité d'effectuer des manipulations génétiques. Ce protocole a fait l'objet d'un article, publié en 2004, intitulé « Expression and rapid purification of highly active hexahistidine-tagged guinea pig liver transglutaminase », qui est présenté au chapitre 4.

Pour cet article, j'ai préparé et séquencé le vecteur d'expression pour le gène de la transglutaminase par biologie moléculaire en plus de participer à l'optimisation de l'expression et de la purification de l'enzyme recombinante avec le Dr Steve Gillet, un chercheur postdoctoral que j'ai formé en lui enseignant les techniques de biologie moléculaire et de biochimie utilisées dans le laboratoire. Le Dr Steve Gillet a effectué toutes les expériences de cinétique enzymatique et a rédigé majoritairement le manuscrit, à l'exception des sections rapportant les méthodes et les résultats de la construction du vecteur pQE32-TGase, dont j'ai été l'auteur.

Article 3.

Expression and rapid purification of highly active hexahistidine-tagged guinea pig liver transglutaminase

Steve M. F. G. Gillet, Roberto A. Chica, Jeffrey W. Keillor and Joelle N. Pelletier

Département de Chimie, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Québec, Canada H3C 3J7

Protein Expression and Purification, 2004, **33**, 256-264

4.1 Abstract

Tissue transglutaminase has been identified as a contributor to a wide variety of diseases, including cataract formation and Celiac disease. Guinea pig tissue transglutaminase has a very broad substrate specificity and therefore is useful for kinetic studies using substrate analogues. Here, we report the expression in *Escherichia coli* of a hexahistidine-tagged guinea pig liver tissue transglutaminase (His_6-tTGase) allowing rapid purification by immobilized-metal affinity chromatography. Using this procedure we have obtained the highest reported specific activity (17 U/mg) combined with a high yield (22 mg/L of culture) for recombinant TGase using a single-step purification protocol. Using two independent spectrophotometric assays, we determined that the K_M value of the recombinant enzyme with the substrate Cbz–Gln–Gly is in the same range as values reported in the literature for the native enzyme. We have thus developed a rapid and reproducible protocol for the preparation of high quality tissue TGase.

Author Keywords: Transglutaminase; Hexahistidine tag; Chaperone; IMAC; Betaine

4.2 Introduction

Transglutaminases (amine γ-glutamyltransferases, EC 2.3.2.13) are a family of enzymes that catalyse Ca^{2+}-dependent acyl transfer reactions. Catalytic activity is exhibited toward the γ-carboxamide moiety of a protein or peptide-bound glutaminyl residue, which serves as acyl donor to either protein or peptide-bound lysyl residues or free amines, resulting in new intermolecular crosslinks [1, 2 and 3]. Tissue transglutaminase (TGase) is widely distributed in animal tissues including liver, kidney, lung, spleen, brain, and the endothelial and smooth muscle cells of arteries, veins, and capillaries [4].

Tissue transglutaminase has been identified as a contributor to cataracts and Celiac disease, and a growing body of evidence suggests that it may be involved in atherosclerosis, inflammation, fibrosis, diabetes, cancer metastases, autoimmune diseases, lamellar ichthyosis, and psoriasis (for a review, see [5]). TGase is also suspected to play a role in neurodegenerative diseases (such as Huntington's disease, Alzheimer's disease, Parkinson's disease, and supranuclear palsy) associated with an increase in polyglutamine-containing peptides in the brain [6, 7 and 8].

Tissue TGase has been purified from many sources, including mammalian livers, rat testicular tissue, and human erythrocytes [9]. Currently, the most widely cited source of native TGase for research is guinea pig liver, due to its rich abundance of endogenous TGase in comparison to other sources of mammalian liver [10]. As much as 25 mg can be obtained from 140 g of guinea pig liver [11], with a specific activity of 18 U/mg, up to 6- to 10-fold higher than the typical commercial native guinea pig tissue TGase (Sigma). As the guinea pig liver TGase shows 88 % sequence homology (80 % sequence identity) at the amino acid level (using standard protein–protein BLAST) with human tissue TGase, its use in medical research is pertinent [12].

The establishment of an efficient expression system of the recombinant mammalian tissue-type transglutaminase would allow the enzyme to be obtained in the high yield and purity that are essential to the structure/function studies required for mechanistic studies and the subsequent development of mechanism-based inhibitors. Since tissue TGases require no post-translational modification for activity, [13] they are a good target for recombinant expression in bacterial cells. Recombinant human endothelial, guinea pig liver and red sea bream (*Pagrus major*) tissue transglutaminases have been expressed and purified [14, 15 and 16]. Expression in *Escherichia coli* and purification of hexahistidine-tagged human tissue transglutaminase have recently been reported by Shi *et al.* [17]. The specific activity reported in the literature for the recombinant human transglutaminases was approximately 2- to 4-fold lower than that of the commercial native guinea pig tissue

TGase (Sigma) and the quantity of enzyme obtained per litre of bacterial culture was in the range of 0.6 mg [17 and 18]. The activity of the recombinant guinea pig liver tissue TGase expressed in *E. coli* obtained by Ikura *et al.* [19] was approximately 7-fold lower than that of the native guinea pig liver tissue TGase, but the quantity of enzyme produced per litre of culture was in the range of 37.5 mg/L. This very high yield indicates the pertinence of pursuing expression in *E. coli*, although improvement of specific activity is essential to our research.

We sought to combine the practical nature of a hexahistidine tag with an improved yield of soluble enzyme to obtain high yields of recombinant guinea pig liver TGase with specific activity that should at least approach that reported for the highly purified native enzyme [11]. Herein, we report two methods for expression in *E. coli* of guinea pig liver tissue transglutaminase with a hexahistidine tag at the N-terminus allowing rapid purification by immobilized-metal affinity chromatography (IMAC).

The first method, based on that of Yokoyama *et al.* [16], relies on the chaperones DnaK and DnaJ to achieve a high yield of soluble protein. The second method relies on the use of betaine, a chemical chaperone, which has been shown to improve the soluble yield during recombinant protein expression [20]. Using our modified procedure we have obtained for the first time a hexahistidine-tagged tissue TGase with the best specific activity (17 U/mg) combined with a high yield (22 mg/L of culture) for recombinant TGase using a single-step purification protocol.

4.3 Materials and methods

4.3.1 Materials

All reagents used were of the highest available purity. Glutamate dehydrogenase (GDH) was purchased from Roche, restriction enzymes and DNA-modifying enzymes were from New England Biolabs and MBI Fermentas. The Sequenase 2.0 DNA sequencing kit as

well as the dITP Nucleotide kit were purchased from Amersham Biosciences. Synthetic oligonucleotides were obtained from AlphaDNA, the infra-red dye labelled sequencing primers were from Li-Cor Biotechnology Division, and QIAEXII gel extraction kit and Ni-NTA agarose resin were obtained from Qiagen. All aqueous solutions were prepared using water purified with a Millipore BioCell system.

4.3.2 Construction of expression plasmid

The guinea pig liver transglutaminase gene was PCR-amplified from plasmid pKTG1 [13] with the primers *Sph*IGTGaseF-Met 5'-(CACACAGCATGCAGGCAG-AGGATCTGATCCTG), which anneals to the 5'-end of the coding region and introduces a unique *Sph*I restriction site (underlined) while removing the start codon, and pKTG1_GTGaseR 5'-(ATCTTCTCTCATCCGCCAAA), annealing ~20 bp downstream from the stop codon and the *Pst*I restriction site, which immediately follows it. The PCR product was digested by *Pst*I/*Sph*I and purified by agarose gel electrophoresis followed by extraction with a QIAEXII kit (Qiagen). The TGase gene was inserted in pQE32 expression vector (Qiagen) via *Sph*I/*Pst*I yielding the plasmid pQE32-TGase. This plasmid encodes an N-terminal His_6-tag (underlined) followed by a five amino acid linker sequence (MRGSHHHHHHGIRMQ) preceding the TGase, resulting in a 704 amino acid protein with a calculated molecular weight of 78.7 kDa. The plasmid was then transformed into *E. coli* XL-1 Blue cells or *E. coli* XL-1 Blue cells containing the plasmid pDnaKJ [16], which encodes the molecular chaperones DnaK and DnaJ. The entire *Sph*I/*Pst*I fragment, including the whole coding region, was verified by DNA sequencing.

4.3.3 Sequence analysis

DNA sequencing was performed using a dITP Nucleotide kit for Sequenase version 2.0 DNA polymerase (Amersham–Pharmacia Biotech) with a Li-Cor IR^2 automated system.

4.3.4 Overexpression of TGase

4.3.4.1 Method A: expression with molecular chaperones DnaK and DnaJ

E. coli, cells from frozen stock, harbouring the expression plasmids pQE32-TGase and pDnaKJ (expressing both DnaK and DnaJ chaperones; [16]) were grown overnight in 50 mL LB medium containing 100 μg/mL ampicillin for maintenance of pQE32-TGase and 30 μg/mL chloramphenicol for maintenance of pDnaKJ. A portion (20 mL) of the bacterial suspension was then inoculated into 400 mL of fresh LB medium and the cells were incubated at 37 °C with agitation (250 rpm). When OD_{600} reached approximately 0.6, the expression of His_6-tTGase was induced by addition of IPTG to a final concentration of 1 mM. The culture was incubated further for 20 h at 28 °C with shaking at 240 rpm.

4.3.4.2 Method B: expression with chemical chaperone betaine

E. coli cells, from frozen stock, harbouring the expression plasmid pQE32-TGase were grown overnight in 50 mL LB medium containing 100 μg/mL ampicillin. A portion (20 mL) of the bacterial suspension was then inoculated into 400 mL of fresh LB medium containing 1 M sorbitol and 2.5 mM betaine and induction of His_6-tTGase expression was undertaken as in Method A.

All of the following procedures were carried out at 0–4 °C. After induction of His_6-tTGase, the bacterial suspension was centrifuged at 1500*g* for 30 min. The cell pellet was resuspended in 8 mL of extract buffer (0.1 M potassium phosphate buffer, pH 8.0, containing 5 mM imidazole) and disrupted by sonication (three pulses of 200 W for 30 s with a tapered micro-tip). The cell debris were removed by centrifugation at 2000*g* for 30 min. The supernatant fraction obtained was used as the cell extract.

4.3.5 Purification of His_6-tTGase

All procedures were carried out at 0–4 °C. To the cell extract, 500 µL Ni–NTA agarose, equilibrated with extract buffer, was added and mixed gently (100 rpm on a rotary shaker) for 60 min. The lysate-Ni–NTA mixture was loaded into a column (Bio-Rad Poly-Prep Chromatography Columns 0.8 × 4 cm). The flow-through was collected and the column was washed successively with 6 mL of extract buffer, 6 mL of washing buffer A (0.1 M potassium phosphate buffer, pH 8.0, containing 10 mM imidazole), and 6 mL of washing buffer B (0.1 M potassium phosphate buffer, pH 8.0, containing 20 mM imidazole). The column was eluted with 4 mL elution buffer (0.1 M potassium phosphate buffer, pH 7.5, containing 250 mM imidazole). The eluant was transferred to an Econo-Pac 10-DG column (Bio-Rad) and the tissue TGase was desalted by elution with 10 mM Mops buffer (pH 7.0). Aliquots of this purified eluant were assayed for tissue TGase activity as described above. Fractions with specific activities higher than 10 U/mg were pooled. The solution was then aliquoted (each fraction containing approximately 500 µg of enzyme) into microcentrifuge tubes and flash-frozen on dry ice for 5 min. The samples were lyophilized using a Labconco Freeze Dry System and stored at −20 °C.

4.3.6 Protein concentrations

Concentrations of total protein were generally determined using the Bio-Rad protein assay, a method based on the Bradford assay [21], using BSA as a standard. Colour development took place over 15 min and absorbance values were measured using a Varian Cary 100 Bio. Standard curves were performed daily. For fractions obtained after elution from the Ni–NTA column, in which TGase was considered to be of high purity, protein concentration was determined using the extinction coefficient for native TGase ($A_{280}^{1\%}$=15.8) [22].

4.3.7 Determination of specific activity

TGase activity was measured by the colorimetric hydroxamate assay procedure using 30 mM benzyloxycarbonyl-L-glutaminylglycine, 1 mM EDTA, 5 mM $CaCl_2$, and 0.1 M hydroxylamine in 0.2 M Tris–acetate at pH 6.0 and 37 °C [23]. The activity assay solution was prepared fresh daily. One unit of enzyme activity (U) is defined as the amount of TGase, which catalyses the formation of 1 μmol of γ-glutamylhydroxamic acid per minute. Specific activity is given as units per milligram of protein.

4.3.8 Electrophoresis

Separation of proteins by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE) was conducted using the Fisher scientific electrophoresis cell according to procedures of Laemmli [24] using minigels (105 × 85 mm dimensions) prepared with 9 % acrylamide resolving gel at pH 8.8, 4 % acrylamide stacking gel at pH 6.8, and 2.7 % crosslinker concentrations. The protein sample (10 μL) was mixed with 10 μL of loading buffer [50 mM Tris–HCl (pH 6.8), 100 mM DTT, 2% w/v SDS, 0.1% w/v bromophenol blue, and 10% v/v glycerol] and boiled for 5 min prior to loading. Protein bands were revealed using 250R Coomassie blue staining technique [25]. Coomassie blue-stained gels were dried. The SDS–PAGE broad-range molecular weight markers (Bio-Rad) were used as standards.

4.3.9 Enzyme purity

The dried gels were scanned using a flatbed document scanner at 400 dpi resolution. The scanned images were analysed using the program Scion Image for Windows (Scion Corporation), which has extensive image analysis capabilities for one-dimensional electrophoretic gels. The background was removed by using the *1D Vertical Subtract Background* command. We used the macros *Gelplot2* to plot the lanes profiles. Each peak

was manually delimited on the baseline with the line drawing tool and then measurements of the areas were taken with the *Automatic Outlining Tool*.

4.3.10 Kinetic assays

4.3.10.1 Method I: DMPDA assay

The kinetic properties of His_6-tTGase were determined using a continuous spectrophotometric assay [26]. All assays were performed in triplicate using six different concentrations of donor substrate ranging from 0.1 to 3 K_M. Nonspecific reactivity was observed above these concentrations. A standard stock buffer solution was prepared by combining 12 mL of 1 M Tris–acetate buffer (pH 7.0), 3 mL of 0.1 M $CaCl_2$, 3 mL of 0.02 M EDTA, and 48 mL of water. To 340 µL of this stock solution were added 20 µL of 20 mM *N*,*N*-dimethyl-1,4-phenylenediamine (DMPDA), 25–120 µL of 5 or 50 mM Cbz-Gln-Gly and the final volume was adjusted to 480 µL through the addition of 0–95 µL of water. The reaction solution was preincubated for 3 min at 37 °C prior to initiation of the enzymatic reaction upon addition of 20 µL (~8µg) of purified His_6-tTGase, giving a final volume of 500 µL. Reaction progress was followed as a linear increase of absorbance measured at 278 nm in a Varian Cary 100 spectrophotometer thermostated at 37 °C. For the blank reactions, 20 µL of water was substituted for TGase enzyme. Slopes were measured over the first <10 % substrate conversion to product and were converted into initial rates by dividing by the extinction coefficient of the authentic reaction product, carbobenzyloxy-L-glutamyl-γ-[4-(dimethylamino)anilido]glycine (8940 ± 55 M^{-1} cm^{-1}) [26].

4.3.10.2 Method II: glutamate dehydrogenase-linked assay (GDH)

Kinetic parameters were determined using a continuous spectrophotometric linked enzyme assay [27]. All assays were performed in triplicate using seven different concentrations of donor substrate ranging from 0.1 to 3 K_M. Nonspecific reactivity was observed above these concentrations. A standard stock buffer solution of 400 mM Mops

buffer (pH 7.0), 10 mM $CaCl_2$, and 2 mM EDTA was prepared. To 250 µL of this stock solution were added 20 µL of 12.5 mM NADH, 20 µL of 250 mM α-ketoglutarate, 20 µL of GDH (2.4 U), 20 µL of 83.5 mM *N*-acetyl lysine methyl ester, 25–120 µL of 5 or 50 mM Cbz–Gln–Gly and the final volume was adjusted to 480 µL through the addition of 30-125 µL of water. The reaction solution was preincubated for 3 min at 37 °C prior to initiation of the enzymatic reaction upon the addition of 20 µL (~8 µg) of His_6-tTGase, giving a final volume of 500 µL. For the corresponding blank solution, the TGase solution was replaced by water. The decrease in absorbance due to the oxidation of NADH was followed against a blank at 340 nm in a Varian Cary 100 spectrophotometer thermostated at 37 °C. After an initial lag, linear slopes of absorbance versus time were measured over the first <10 % of substrate conversion to product and were converted into initial rates by dividing by the extinction coefficient of NADH (6220 M^{-1} cm^{-1}) [27].

4.4 Results and discussion

4.4.1 DNA sequencing

DNA sequencing using the automated Li-Cor system revealed a silent mutation at residue Arg76 (CG<u>C</u> instead of CGG) relative to the published sequence of the guinea pig liver TGase gene [28]. Other mutations relative to the published sequence [28] were also detected. A two base pair inversion (TC to CT) in adjacent nucleotides from codons encoding residues Pro141 and Arg142 was observed resulting in a silent mutation at the codon encoding Pro141 (CC<u>C</u> instead of CCT) and a missense mutation at the codon encoding Arg142 (<u>T</u>GG instead of CGG), which results in a Arg to Trp mutation. Trp may be the correct residue at this position, considering that all eukaryotic TGases that have been sequenced encode a Trp residue at this position. A second inversion of two base pairs was detected in the codon encoding residue Arg436 (C<u>AG</u> instead of CGA), which results in a Arg to Gln mutation. While Arg436 exists in other sequences, Gln is found in TGase 1

from *Canis familiaris* at the corresponding position. It is unclear if the Gln436 sequenced here represents the correct native residue, a mutation or a polymorphism.

4.4.2 Expression and purification of His_6-tTGase

Since large amounts of highly active tissue TGase are required for structure/activity studies or for mutation of the enzyme, the establishment of an efficient expression system of the recombinant mammalian tissue-type TGase was needed. We report the development of a rapid and easy one-step purification protocol for the reproducible preparation of tissue TGase with a good yield and high purity. Furthermore, the presence of the hexahistidine tag permits the use of a common purification protocol should mutated enzymes be produced.

The *E. coli* cells harbouring the expression plasmid pQE32-TGase were initially tested for production of the His_6-tagged guinea pig liver transglutaminase in the absence of additional factors promoting protein folding (Figure 1, lanes 1–3). SDS–PAGE analysis of the supernatant and cell pellet following cell lysis by sonication showed a bold protein band migrating slightly more slowly than the native guinea pig liver TGase (as a result of the additional hexahistidine tag (results not shown)). While the transglutaminase band was detected in the soluble fraction, the majority of the expressed enzyme was present in the insoluble fraction. First, we attempted to convert the inclusion bodies to soluble, active enzyme, beginning with denaturation of the inclusion bodies in urea or guanidinium chloride. However, all attempts at refolding by immobilized metal-ion affinity chromatography (Ni–NTA) while applying a decreasing gradient of guanidinium chloride, urea or pH [29 and 30] with or without 5% polyethylene glycol [19] failed to regenerate active enzyme from the inclusion bodies (data not shown).

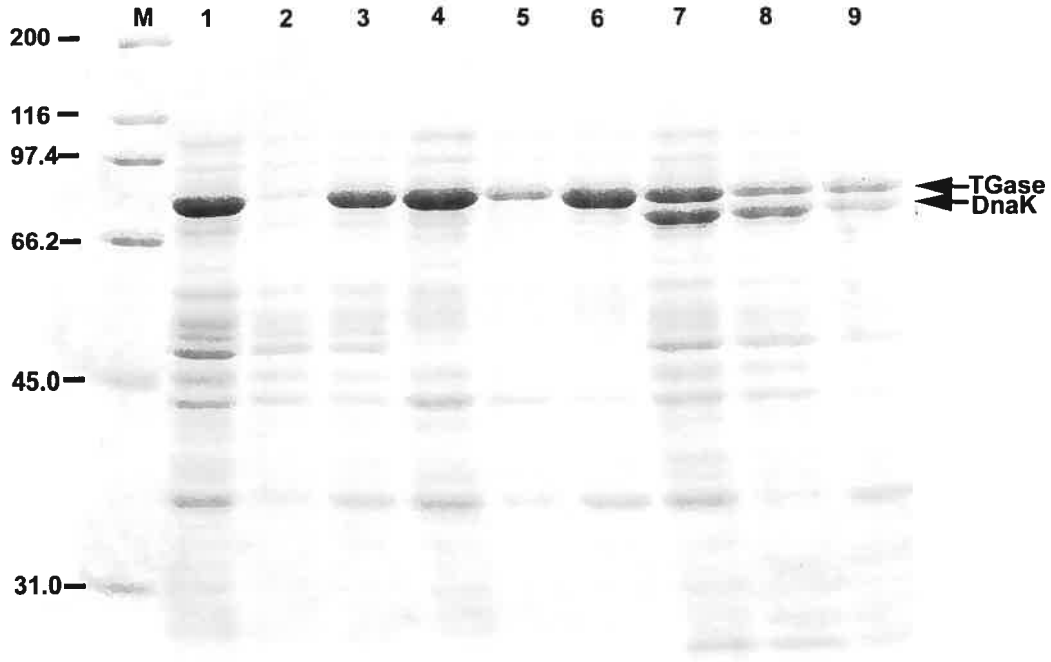


Figure 1. TGase expression in the presence or absence of folding modulators. SDS–PAGE patterns of total lysate, soluble, and insoluble fractions from *E. coli* cells expressing TGase are presented. Five microlitres of each fraction was applied. Lane M, prestained protein standard. Lanes 1–3, *E. coli* cells harbouring pQE32-TGase plasmid induced in the absence of additional factors promoting protein folding; lane 1, total lysate; lane 2, soluble fraction; and lane 3, insoluble fraction. Lanes 4–6, *E. coli* cells harbouring pQE32-TGase plasmid induced in the presence of chemical chaperone (sorbitol/betaine); lane 4, total lysate; lane 5, soluble fraction; and lane 6, insoluble fraction. Lanes 7–9, *E. coli* cells harbouring pQE32-TGase plasmid induced in the presence of molecular chaperones (DnaK and DnaJ); lane 7, total lysate; lane 8, soluble fraction; and lane 9, insoluble fraction. The arrows indicate His_6-tTGase (TGase) and the molecular chaperone DnaK.

Recently, it has been reported that co-overexpression of chaperones has a positive effect on the production of soluble and active recombinant transglutaminase [19]. The results of experiments in which DnaK and DnaJ were co-overexpressed with His_6-tTGase

(Method A) showed the solubilizing effect of these molecular chaperones (Figure 1, lanes 7-9). Different concentrations of IPTG (2, 20, 100 μM, 1 and 5 mM; 20 h induction) were compared for their effectiveness in inducing maximal expression of soluble active transglutaminase in the presence of these chaperones. Little expression was observed after induction at 2 μM IPTG. Soluble expression was similar at all other concentrations, though increased somewhat at concentrations ≥1 mM, with no further increase at 5 mM (data not shown).

The effect of incubation temperature on the yield of soluble, active His_6-tTGase in the presence of the molecular chaperones was also examined (Figure 2). The specific activity of the crude cell extract was maximal when the induction was performed at 28 °C, notwithstanding the total production level of transglutaminase that appears to be relatively constant from 20 to 37 °C (Figure 2). This suggests that proper folding of tissue TGase into its active form is maximal at 28 °C. Furthermore, the amount of TGase in the soluble fraction decreased markedly as the temperature rose above 32 °C (Figure 2).

The results showed that the co-overexpression of the molecular chaperones DnaK and DnaJ (Method A) enhances production of soluble, active protein. The soluble His_6-tTGase was purified by IMAC as described under Materials and methods. The yield of soluble TGase per litre of culture obtained after purification was determined for each induction temperature and shows no significant differences, except for the decrease at 37 °C (Figure 3). Increased *E. coli* cell density after overnight induction at 32 °C relative to 28 °C partially compensates for the lower percent solubility, giving a similar yield (Figure 3). In keeping with these results, all further expression was undertaken at 28 °C.

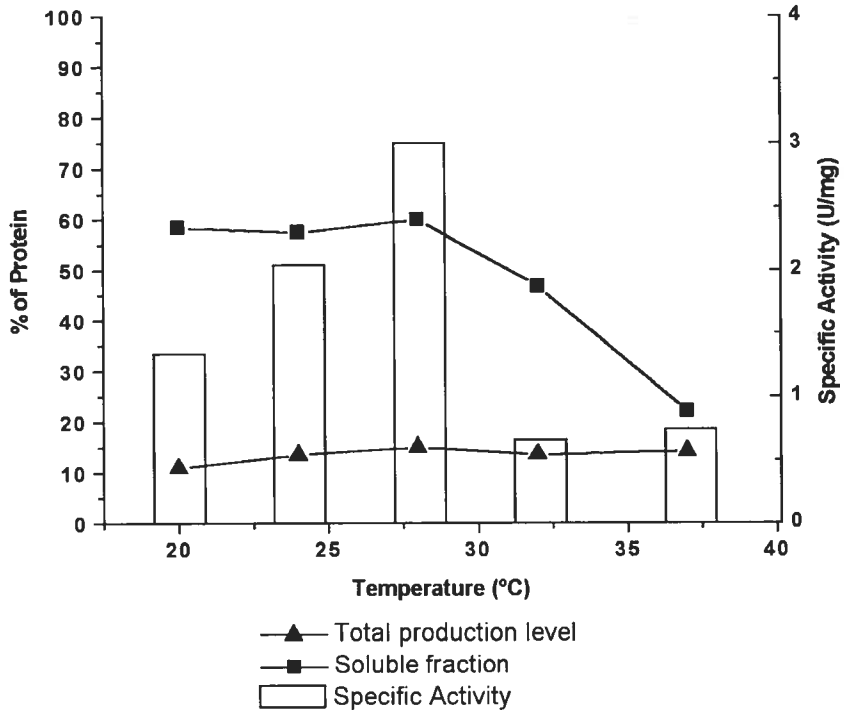


Figure 2. Specific activity of transglutaminase in the soluble fractions obtained from lysates of *E. coli* expressing His_6-tTGase, DnaK, and DnaJ, induced at different temperatures. Filled squares (■) indicate the portion of transglutaminase in the soluble fraction at each temperature. Filled triangles (▲) indicate the total production level of transglutaminase in the cells at each temperature (% of the total protein production).

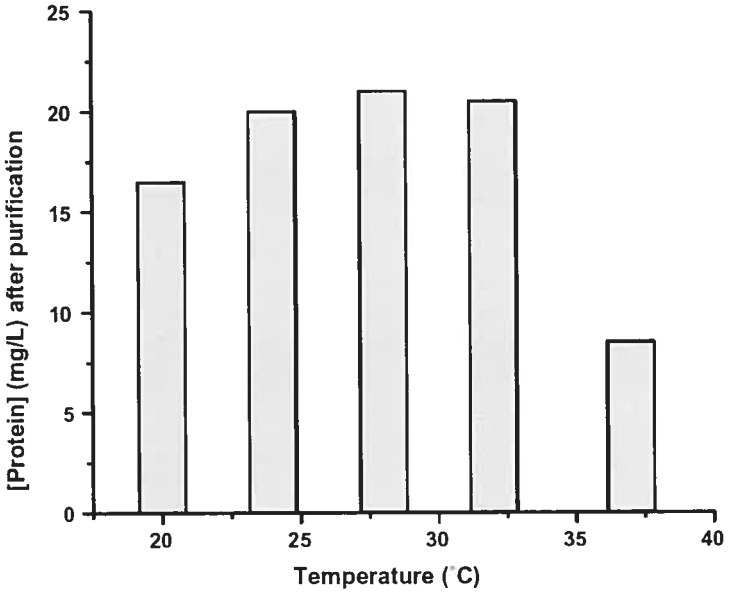


Figure 3. Quantity of His_6-tTGase per litre of culture obtained after purification, for different induction temperatures.

Upon evaluation of different washing and elution protocols, it was determined that His_6-tTGase binds fairly weakly to Ni–NTA. When loading was attempted using buffer containing 10 mM imidazole, His_6-tTGase did not bind to a significant extent. After binding was achieved in "extract buffer" (containing 5 mM imidazole), we observed that washing buffers containing concentrations of imidazole higher than 20 mM were able to effect the dissociation of most of the tissue TGase from the Ni–NTA resin. SDS–PAGE analysis of the purification steps is shown in Figure 4A. A summary of a typical purification procedure is shown in Table 1 (Method A). Based on those results, the production of the soluble recombinant transglutaminase was calculated to be 22 mg/L culture with a specific activity of 17 U/mg and a purity of 84 % determined by densitometry. By comparison, 25 mg of native transglutaminase can be obtained from 140 g of guinea pig liver with a specific activity of 18 U/mg and a purity around 84 % in the best conditions [11] and the typical commercial native guinea pig tissue TGase (Sigma)

has a specific activity of 2-3 U/mg. Two principal impurities in the elution fraction were observed by SDS–PAGE, one migrating at approximately 70 kDa (Figure 4A: immediately below His_6-tTGase) and the second at approximately 66 kDa. In Figure 5, the elution fraction of the control purification of cells expressing DnaK and DnaJ but not the His_6-tTGase (lane 2) shows the same band that migrates near 70 kDa. This is apparently due to non-specific binding of DnaK (M_r 72,000) on the Ni–NTA resin as evidenced by the absence of this impurity in the elution fraction of cells lacking pDnaKJ plasmid (lanes 3 and 4).

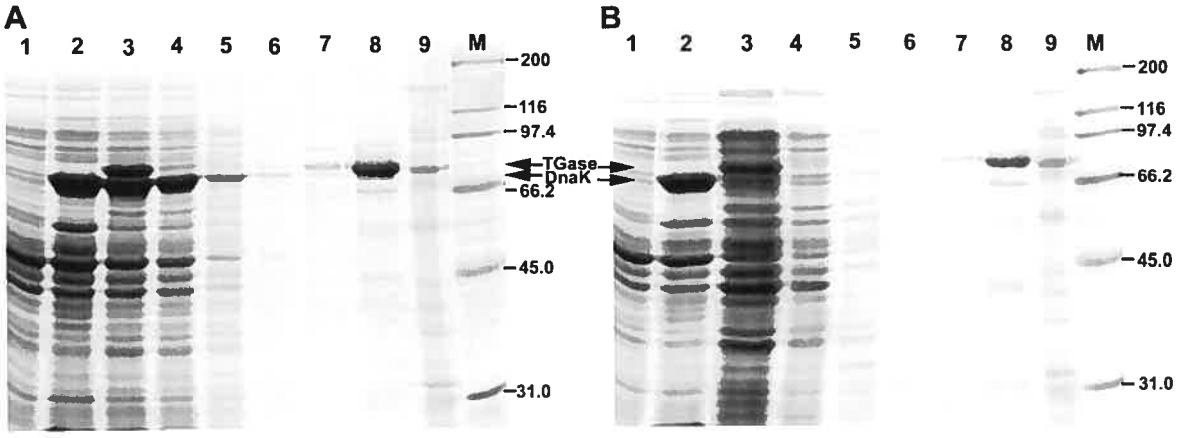


Figure 4. SDS–PAGE of fractions obtained during the purification. Expression with molecular chaperones DnaK (70 kDa) and DnaJ (40 kDa) (A) or with chemical chaperone (betaine) (B). Ten microlitres of each fraction was applied. For both panels, lane M, prestained protein standards; lane 1, negative control: cell lysate from *E. coli* without expression plasmid; lane 2, negative control: cell lysate from *E. coli* harbouring pDnaKJ plasmid alone; lane 3, cell lysate from either expression system; lane 4, flow-through of the Ni–NTA; lane 5, first washing fraction (5 mM imidazole); lane 6, second washing fraction (10 mM imidazole); lane 7, third washing fraction (20 mM imidazole); lane 8, elution fraction (250 mM imidazole); and lane 9, native guinea pig liver transglutaminase purified according to the procedure of Leblanc *et al.* [11]. The arrows indicate His_6-tTGase (TGase) and the molecular chaperone DnaK.

Table 1. Activity of recombinant guinea pig liver His_6-tTGase from the two *E. coli* expression systems

Expression system	Fraction	Total protein (mg)	Total activity (U[a])	Specific activity (U/mg)	Yield (%)[b]	Purity (%)
Method A. (with molecular chaperones)	Cell extract[c]	14.0	42	3.0	100	N.D.
	Purified enzyme	1.1	18.7	17.0	44.5	84
Method B. (with chemical chaperone)	Cell extract[c]	16.6	24.9	1.5	100	N.D.
	Purified enzyme	0.7	3.9	5.5	15.7	80

[a] Units are μmol of γ-glutamylhydroxamic acid formed per minute

[b] Total activity of the purified enzyme relative to the total activity of the cell extract

[c] Cell extract (3mL) prepared from 50 mL of culture was used as a starting material

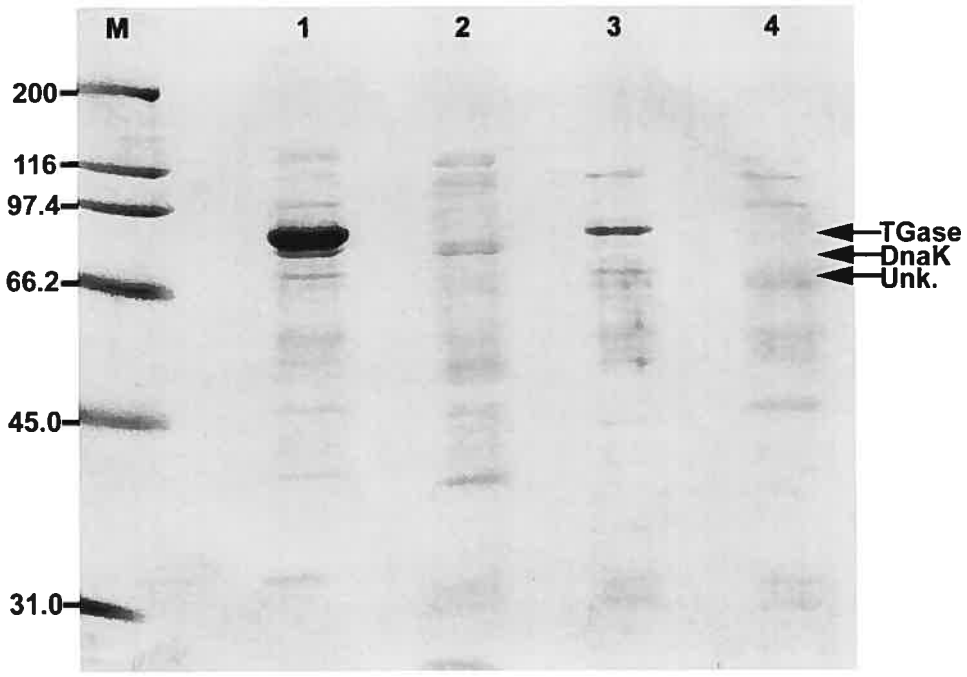


Figure 5. SDS–PAGE of fractions obtained after purification. Ten microlitres of each fraction was applied. Lane M, prestained protein standards; lane 1, elution fraction from *E. coli* harbouring plasmids pDnaKJ and pQE32-TGase; lane 2, elution fraction from *E. coli* harbouring pDnaKJ alone; lane 3, elution fraction from *E. coli* harbouring pQE32-TGase alone; and lane 4, *E. coli* without expression plasmid. The arrows indicate His_6-tTGase (TGase), the molecular chaperone DnaK, and the putative C-terminal tissue TGase degradation product (Unknown).

To avoid the purification problems associated with co-elution of chaperone, we undertook a second method of expression involving the use of osmotic stress to facilitate the uptake of betaine (Method B). Betaine is believed to be excluded from the immediate domains of proteins, causing thermodynamically unfavourable 'preferential hydration' and, thus, minimization of solvent–protein contact and stabilization of protein structure [20]. We propagated *E. coli* in the presence of 2.5 mM betaine in 50 mL culture and induced tissue

TGase production at 28 °C for 20 h in 1 M sorbitol LB-based medium (Method B). Sorbitol/betaine (S/B) treatment enhances production of soluble, active protein relative to expression of His_6-tTGase in the absence of any chaperone (Figure 1, lanes 4–6). The soluble His_6-tTGase was purified by IMAC as described under Materials and methods. SDS–PAGE analysis of the purification steps is shown in Figure 4B. Table 1 shows a summary of a typical purification procedure. Based on these results, the production of the soluble recombinant transglutaminase was calculated to be 14 mg/L culture with a specific activity of 5.5 U/mg and a purity of 80 %. While this strategy eliminated the impurity around 70 kDa, the band migrating at approximately 66 kDa was still present. This impurity was absent from the elution fraction of the purification from cells not expressing the His_6-tTGase (Figure 5, lanes 2 and 4). This band may be due to the partial C-terminal degradation of the recombinant tissue TGase. While the use of chemical chaperones [20] increased the yield of soluble enzyme relative to expression in the absence of molecular chaperones, the specific activity of the resulting purified tissue TGase was lower than that obtained by Method A. As we required enzyme of the highest possible specific activity for our research needs, we did not further investigate Method B, although it provides enzyme of reasonable quality with the potential advantage of carrying no trace of contaminating DnaK chaperone.

Expression of His_6-tTGase in *E. coli* by combination of Methods A and B (with DnaK, DnaJ, and sorbitol/betaine medium) did not show any enhancement of solubility relative to Method A (data not shown).

4.4.3 Kinetic of recombinant guinea pig liver His_6-tTGase

The specific activity of purified His_6-tTGase was determined according to the hydroxamate assay using Cbz–Gln–Gly and hydroxylamine [23]. This value (17 U/mg) was similar to that of the highly purified native transglutaminase (18 U/mg [11] or 14 U/mg [23]) previously determined using the same method. Using the DMPDA assay (Method I)

[26] we determined K_M and k_{cat} for the well-characterized donor substrate Cbz–Gln–Gly. Nonlinear regression to the Michaelis–Menten equation yielded a K_M for Cbz–Gln–Gly as 5 ± 2 mM and a k_{cat} as 60 min^{-1}. Using the GDH-linked assay (Method II), with the same donor substrate, we determined a K_M of 3.9 ± 0.4 mM and a k_{cat} of 44 min^{-1}. The K_M value reported in the literature for the native guinea pig tissue TGase for this donor substrate using Method I is 2.40 ± 0.05 mM [26] and using Method II is 5.7 ± 0.6 mM [31], which are comparable to the values we have obtained. In the literature, the k_{cat} value reported for the native guinea pig tissue TGase for Cbz–Gln–Gly using Method II is 52 ± 2 min^{-1} [31]. This indicates that the hexahistidine tag and the potential mutations or polymorphisms at positions 142 and 436 detected by automated DNA sequencing do not appear to affect the catalytic activity of the recombinant enzyme since the k_{cat} and K_M values are not significantly different than those of the tissue TGase purified from guinea pig liver.

4.5 Future work

The yield, purity, and specific activity of the TGase purified by Method A are the highest reported to date using a single-step purification protocol, indicating that this rapid method yields enzyme of high enough quality for most research applications. Its kinetic properties do not significantly differ from those of the native enzyme. As detailed above, it seems probable that the two principal impurities are the chaperone DnaK and a C-terminal degradation product of the tissue TGase. DnaK could be rapidly eliminated by adding a FLAG tag (DYKDDDDK) that would allow elimination of the tagged chaperone by passing the sample over an anti-FLAG affinity column [32] or by adding any other such affinity label. One could also consider using other chaperones than DnaK and DnaJ. The putative C-terminal tissue TGase degradation product could be eliminated by transferring the hexahistidine tag from the N- to the C-terminus such that C-terminally degraded tissue TGase will not bind to Ni–NTA. Transferring the hexahistidine tag to the C-terminus may also have the added benefit of improving the affinity of His_6-tTGase toward Ni–NTA,

considering the relatively poor affinity we observed for the N-terminally tagged tissue TGase (Figure 4).

4.6 Conclusions

In conclusion, the procedures for the expression and purification of recombinant guinea pig liver transglutaminase presented in this work represent an efficient, easily reproducible, and rapid method to obtain highly purified enzyme with high specific activity. While the use of sorbitol and betaine as chemical chaperones [20] increased the yield of soluble enzyme relative to expression in the absence of molecular chaperones, the specific activity of the resulting purified tissue TGase was not high enough to warrant further investigation. Expression in the presence of the molecular chaperone DnaK and DnaJ [15], coupled to rapid affinity purification, resulted in a good yield of highly active enzyme.

4.7 Acknowledgements

We thank Dr. H. Yasueda (Central Research Laboratories, Ajinomoto Co. Inc., Japan) for pDnaKJ plasmid, Dr. K. Ikura (Department Of Applied Biology, Kyoto Institute Technology, Kyoto, Japan) for the pKTG1 plasmid, C. Gravel (Université de Montréal) for the native enzyme, N. Doucet (Université de Montréal) for critical reading of the manuscript, and D. Halim (Université de Montréal) for assistance with the kinetic assays. This work was supported by VRQ Grant 2201-147 and FQRNT Grant ER-80357, jointly held by J.W.K. and J.N.P. R.A.C. is also grateful to FQRNT.

4.8 References

1. J.E. Folk and P.W. Cole, Transglutaminases: mechanistic features of the active-site as determined by kinetic and inhibitor studies. *Biochim. Biophys. Acta* **122** (1996), pp. 244–264.

2. K.E. Achyuthan, T.F. Slaughter, M.A. Santiago, J.J. Enghild and C.A. Greenberg, Factor XIIIa-derived peptides inhibit transglutaminase activity. Localization of substrate recognition sites. *J. Biol. Chem.* **268** (1993), pp. 21284–21292.

3. C.S. Greenberg, P.J. Birkbichler and R.H. Rice, Transglutaminases: multifunctional cross-linking enzymes that stabilize tissues. *FASEB J.* **5** (1991), pp. 3071–3077.

4. D. Aeschlimann and M. Paulsson, Transglutaminases: protein cross-linking enzymes in tissues and body fluids. *Thromb. Haemostasis* **71** (1994), pp. 402–415.

5. S.-Y. Kim, T.M. Jeitner and P.M. Steinert, Transglutaminases in disease. *Neurochem. Int.* **40** 1 (2002), pp. 85–103.

6. A.J.L. Cooper, T.M. Jeitner, V. Gentile and J.P. Blass, Cross linking of polyglutamine domains catalyzed by tissue transglutaminase is greatly favored with pathological-length repeats: does transglutaminase play a role in $(CAG)_n/Q_n$-expansion disease?. *Neurochem. Int.* **40** 1 (2002), pp. 53–67.

7. S.M. Singer, G.M. Zainelli, M.A. Norlund, J.M. Lee and N.A. Muma, Transglutaminase bonds in neurofibrillary tangles and paired helical filament tau early in Alzheimer's disease. *Neurochem. Int.* **40** 1 (2002), pp. 17–30.

8. M.V. Karpuj, M.W. Becker and L. Steinman, Evidence for a role for transglutaminase in Huntington's disease and the potential therapeutic implications. *Neurochem. Int.* **40** 1 (2002), pp. 31–36.

9. B. Wilhelm, A. Meinhardt and J. Seitz, Transglutaminases: purification and activity assays. *J. Chromatogr. B* **684** (1996), pp. 163–177.

10. T. Abe, S.I. Chung, R.P. DiAugustine and J.E. Folk, Rabbit liver transglutaminase: physical, chemical and catalytic properties. *Biochemistry* **16** (1977), pp. 5495–5501.

11. A. Leblanc, N. Day, A. Ménard and J.W. Keillor, Guinea pig liver transglutaminase: a modified purification procedure affording enzyme with superior activity in greater yield. *Protein Expr. Purif.* **17** (1999), pp. 89–95.

12. W. Dietrich, T. Ehnis, M. Bauer, P. Donner, U. Volta, E.O. Riecken and D. Schuppan, Identification of tissue transglutaminase as the autoantigen of celiac disease. *Nat. Med.* **3** (1997), pp. 797–801.

13. K. Ikura, Y. Tsuchiya, R. Sasaki and H. Chiba, Expression of guinea-pig liver transglutaminase cDNA in *Escherichia coli*. Amino-terminal *N*-alpha-acetyl group is not essential for catalytic function of transglutaminase. *Eur. J. Biochem.* **187** (1990), pp. 705–711.

14. K.N. Lee, S.A. Arnold, P.J. Birckbilcher, M.K. Patterson, Jr., B.M. Fraij, Y. Takeuchi and H.A. Carter, Site-directed mutagenesis of human tissue transglutaminase: Cys-277 is essential for transglutaminase activity but not for GTPase activity. *Biochim. Biophys. Acta* **1202** (1993), pp. 1–6.

15. K. Ikura, Y. Tsuchiya, R. Sasaki and H. Chiba, Expression of guinea-pig liver transglutaminase cDNA in *Escherichia coli*. Amino-terminal *N*-α-acetyl group is not essential for catalytic function of transglutaminase. *Eur. J. Biochem.* **187** (1990), pp. 705–711.

16. K.-I. Yokoyama, Y. Kikuchi and H. Yasueda, Overproduction of DnaJ in *Escherichia coli* improves in vivo solubility of the recombinant fish-derived transglutaminase. *Biosci. Biotechnol. Biochem.* **62** 6 (1998), pp. 1205–1210.

17. Q. Shi, S.-Y. Kim, J.P. Blass and A.J.L. Cooper, Expression in *Escherichia coli* and purification of hexahistidine-tagged human tissue transglutaminase. *Protein Expr. Purif.* **24** (2002), pp. 366–373.

18. A. Ambrus and L. Fésüs, Polyethylene glycol enhanced refolding of the recombinant human tissue transglutaminase. *Prep. Biochem. Biotechnol.* **31** 1 (2001), pp. 59–70.

19. K. Ikura, T. Kokubu, S. Natsuka, A. Ichikawa, M. Adachi, K. Nishiara, H. Yanagi and S. Utsumi, Co-overexpression of folding modulators improves the solubility of the recombinant guinea pig liver transglutaminase expressed in *Escherichia coli*. *Prep. Biochem. Biotechnol.* **32** 2 (2002), pp. 189–205.

20. J.R. Blackwell and R. Horgan, A novel strategy for production of a highly expressed recombinant protein in an active form. *FEBS Lett.* **295** 1–3 (1991), pp. 10–12.

21. M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal. Biochem.* **72** (1976), pp. 248–254.

22. J.E. Folk and S.I. Chung, Transglutaminases. *Methods Enzymol.* **113** (1985), pp. 358–364.

23. J.E. Folk and P.W. Cole, Mechanism of action of guinea pig liver transglutaminase. I. Purification and properties of the enzyme: identification of a functional cysteine essential for activity. *J. Biol. Chem.* **241** (1966), pp. 5518–5525.

24. U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227** (1970), pp. 680–685.

25. C. Fernandez-Patron, L. Castellanos-Serra and P. Rodriguez, Reverse staining of sodium dodecyl sulfate polyacrylamide gels by imidazole-zinc salts: sensitive detection of unmodified proteins. *Biotechniques* **12** (1992), pp. 564–573.

26. P. de Macédo, C. Marrano and J.W. Keillor, A direct continuous spectrophotometric assay for transglutaminase activity. *Anal. Biochem.* **285** (2000), pp. 16–20.

27. N. Day and J.W. Keillor, A continuous spectrophotometric linked enzyme assay for transglutaminase activity. *Anal. Biochem.* **274** (1999), pp. 141–144.

28. K. Ikura, T. Nasu, H. Yokota, Y. Tsuchiya, R. Sasaki and H. Chiba, Amino acid sequence of guinea pig liver transglutaminase from its cDNA sequence. *Biochemistry* **27** 8 (1988), pp. 2898–2905.

29. C. Panagabko, S. Morley, S. Neely, H. Lei, D. Manor and J. Atkinson, Expression and refolding of recombinant human α-tocopherol transfer protein capable of specific α-tocopherol binding. *Protein Expr. Purif.* **24** (2002), pp. 395–403.

30. Z. Gu, M. Weidenhaupt, N. Ivanova, M. Pavlov, B. Xu, Z.-G. Su and J.-C. Janson, Chromatographic methods for the isolation of, and refolding of proteins from, *Escherichia coli* inclusion bodies. *Protein Expr. Purif.* **25** (2002), pp. 174–179.

31. N. Day, Études mécanistiques de l'étape d'acylation de la transglutaminase tissulaire de foie de cobaye, Mémoire de thèse, Université de Montréal, 1999

32. T.P. Hopp, K.S. Prickett, V.L. Price, R.T. Libby, C.J. March, D.P. Cerretti, D.L. Urdal and P.J. Conlon, A short polypeptide marker sequence useful for recombinant protein identification and purification. *Biotechnology* **6** (1988), pp. 1204–1210.

CHAPITRE 5

Développement d'un modèle d'homologie de la TGase de foie de cobaye et mutagenèse de résidus du site actif

5.0 Préface

Pour pouvoir entreprendre l'évolution dirigée de la TGase de foie de cobaye par une approche semi-aléatoire, nous avions besoin d'information structurelle et fonctionnelle additionelle. Dans ce chapitre, nous présentons le premier modèle par homologie de la TGase de foie de cobaye, qui nous a permis d'acquérir de l'information structurelle supplémentaire. De plus, de l'information fonctionnelle a été acquise lors de cette étude par des expériences de mutagenèse dirigée et combinatoire de certains résidus du site actif de la TGase de foie de cobaye, dont le rôle est mal défini dans la littérature. Cette information a été d'une grande utilité pour l'application de notre approche semi-aléatoire.

Pour le manuscrit « Homology modeling of guinea pig liver transglutaminase and mutagenesis of conserved active-site residues Tyr519 and Cys336 », présenté au chapitre 5, j'ai créé le modèle d'homologie de la TGase de foie de cobaye, j'ai construit le vecteur d'expression pQE32-GTG et j'ai réalisé toutes les expériences de cinétique enzymatique et la majorité des expériences de mutagenèse dirigée, en plus de rédiger le manuscrit. Le Dr Steve Gillet a développé la méthode de cinétique enzymatique à la 7-hydroxycoumarine en plus de synthétiser le composé Cbz-Gly-7HC. Farah-Jade Dryburgh m'a assisté lors de la création des mutants Cys336 et Tyr519 de la TGase de foie de cobaye durant son stage sous ma supervision. Ce manuscrit était soumis pour publication au moment du dépôt de cette thèse.

Article 4.

Homology modeling of guinea pig liver transglutaminase and mutagenesis of conserved active-site residues Tyr519 and Cys336

Roberto A. Chica, Steve M.F.G. Gillet, Farah-Jade Dryburgh, Jeffrey W. Keillor and Joelle N Pelletier

Département de chimie, Université de Montréal, CP 6128, Succursale Centre-Ville, Montréal, Québec, H3C 3J7, Canada

Manuscrit soumis

5.1 Abstract

Transglutaminases (TGases) cross-link peptides and proteins by catalyzing the formation of γ-glutamyl-ε-lysyl bonds. Crystal structures from TGases all show a conserved Tyr within hydrogen bonding distance of the catalytic Cys. In tissue TGases, the catalytic Cys is located between this conserved Tyr and a second conserved Cys. It has been proposed that removal of this conserved Tyr would lead to immediate inactivation of the enzyme as the catalytic Cys would form a disulphide bridge with this second Cys residue. In addition, our previous results from molecular modeling suggested that this Tyr forms a critical H-bond with the acyl-donor substrate. To investigate these hypotheses, we turned to the widely-studied guinea pig liver TGase. Because its structure has not been reported, we generated a homology model for this enzyme. Owing to its 43 % sequence identity with red sea bream tissue TGase, the model exhibits high structural similarity with this and other tissue TGases, justifying the use of guinea pig liver TGase as our experimental system. Mutagenesis of the conserved Tyr519 and Cys336 resulted in only a two- to four-fold decrease in catalytic efficiency, indicating that these two residues are not essential for activity. Furthermore, wild-type and Cys336 mutants lose activity upon incubation at 37 °C. Activity was not restored by addition of a reducing agent, invalidating the hypothesis that residue Cys336 forms an inactivating disulphide bond with the catalytic Cys. Our work provides the first model of guinea pig liver TGase and demonstrates that its active-site is surprisingly accommodating of mutations.

Keywords: tissue transglutaminase, TGase, homology modeling, site-directed mutagenesis, combinatorial mutagenesis, enzyme kinetics

5.2 Introduction

Eukaryotic transglutaminases (protein-glutamine γ-glutamyltransferases, EC 2.3.2.13) are a family of enzymes that catalyze the cross-linking of peptides and proteins by

Ca^{2+}-dependent acyl transfer reactions [1-5]. They catalyze the formation of isopeptide bonds between the γ-carboxamide group of a protein- or peptide-bound glutamine side chain and the ε-amino group of a protein- or peptide-bound lysine side chain *via* a modified ping-pong mechanism [6]. The reaction is carried out by a conserved Cys-His-Asp catalytic triad that is similar to that of cysteine proteases: residues Cys and His are believed to form a thiolate-imidazolium ion pair. The first step of the mechanism is the acylation step where a glutamine-containing protein or peptide acting as an acyl-donor reacts with the enzyme's catalytic cysteine thiolate, releasing NH_3 and generating the covalent acyl-enzyme intermediate. This intermediate then reacts with a second substrate, the acyl-acceptor, which can be almost any primary amine [7], to yield the amide product and free enzyme in the second step of the TGase mechanism, the deacylation step. The acyl-enzyme intermediate can also undergo hydrolysis in the absence of primary amine but at a slower rate. It has been shown that the deacylation step of the TGase mechanism is the rate-limiting step [8, 9] even when the acceptor substrate is highly nucleophilic, demonstrating the efficiency with which TGases activate the acyl-transfer from the glutamine γ-carboxamide group.

Eukaryotic TGases are divided into nine classes [5]. Thus far, crystal structures from only four TGases belonging to three different classes have been determined: plasma TGase (factor XIIIa) [10-13], tissue TGase [14, 15] and epidermal TGase [16-19]. Plasma TGase is a homodimeric protein that requires cleavage of a propeptide by thrombin for activation. Tissue TGase and epidermal TGase are intracellular, monomeric enzymes that are inhibited by GTP [14, 17]. Epidermal TGase also requires proteolytic activation for activity, unlike tissue TGase. All eukaryotic TGases share a common structure that consists of four domains: the N-terminal β-sandwich domain, the catalytic core which contains the conserved catalytic triad, the barrel 1 domain and the C-terminal barrel 2 domain.

The crystal structures of human factor XIIIa [12], red sea bream tissue TGase [15], human tissue TGase [14] and human epidermal TGase [16] all show a conserved Tyr

residue that is within hydrogen bonding distance of the catalytic Cys residue. It has been proposed that prior to catalysis, the putative hydrogen bond between the conserved Tyr and the catalytic Cys must be broken in order to give the glutamine-containing substrate access to the catalytic thiolate [12]. However, our previous molecular modeling study performed on red sea bream tissue TGase suggested that this Tyr residue instead forms a critical H-bond with the γ-carboxamide group of the acyl-donor substrate, orienting and activating it for the nucleophilic attack in the acylation step of the TGase mechanism [20]. In tissue TGases, the catalytic Cys is located between this conserved Tyr and a second, conserved Cys. It has also been proposed that removal of the conserved Tyr of red sea bream tissue TGase would lead to immediate inactivation as the catalytic Cys would form a disulphide bridge with this second Cys [15]. Such a disulphide bridge has not been observed but indirect observations made by Folk and coworkers [21, 22] support this hypothesis. Namely, they reported that another tissue TGase, guinea pig liver TGase, loses activity steadily during storage; much of the lost activity can be restored by addition of the reducing agent dithiothreitol, consistent with inactivation having been the result of disulphide bridging of the catalytic Cys.

To investigate these hypotheses, a homology model for guinea pig liver TGase was developed since no crystal structure of this enzyme has been reported. This model demonstrated that guinea pig liver TGase has an active-site geometry and overall structure that is similar to that of red sea bream tissue TGase from which the functional hypotheses were derived. Thus, we generated mutants of the most extensively studied TGase, guinea pig liver TGase, at the positions of the conserved active-site Tyr (Tyr519) and the second Cys residue (Cys336). Cys336 was mutated to Ala and Val to verify if it participates in oxidation of the catalytic residue Cys277 by formation of a disulphide bond. Tyr519 was mutated to Leu, Phe, Gln or His to modify the putative H-bond with the donor substrate and to verify if removal of this residue inactivates the enzyme by allowing disulphide bond formation with the catalytic Cys277.

5.3 Materials and methods

5.3.1 Homology modeling

The homology model of guinea pig liver TGase was prepared, using as templates the PDB coordinates of the crystal structures of human tissue TGase (1KV3), red sea bream tissue TGase (1G0D), human epidermal TGase (1L9M) and human factor XIIIa (1EVU). Crystallized water molecules, heterologous molecules and extra monomers were removed. The primary sequences of these TGases were extracted from their crystal structures using the Homology module of InsightII (Accelrys, San Diego, CA). Using these sequences, an alignment was derived using either the Tcoffee server [23] or ClustalW [24], with default parameters. Five 3D models for guinea pig liver TGase were derived for each sequence alignment file using the crystal structure coordinates by Modeller 8v1 [25] with default parameters for generation of loop conformations. The quality of the proposed models was assessed using the Molprobity [26] web server, Whatcheck [27] and ProsaII [28]. All ten models were then refined using a series of energy minimizations performed with the Discover module of InsightII. We applied a tether restraint of 100 kcal/Å^2 on all atoms using 1000 steps of steepest descents energy minimization followed by a conjugate gradients energy minimization until convergence of 0.001 kcal mol^{-1}Å^{-1} with a distance-dependent dielectric constant of 4 to mimic the interior of proteins. The resulting structures were again analyzed using Molprobity, Whatcheck and ProsaII. On the basis of those results, model T2 was selected as the representative model. It was subjected to an unconstrained molecular dynamics simulation consisting of a 1 ps equilibration of the molecular system at 300 K followed by the actual simulation to explore conformational space for 300 ps while maintaining the same temperature. A nonbond cutoff of 15 Å was used to speed up calculations. The radius of gyration of the structure during the simulation was measured using the Decipher module of InsightII.

5.3.2 Materials

All reagents used were of the highest available purity. Restriction enzymes and DNA-modifying enzymes were from New England Biolabs (Mississauga, ON) and MBI Fermentas (Burlington, ON). Synthetic oligonucleotides were obtained from Alpha DNA (Montreal, QC) and Integrated DNA Technologies (Coralville, IA), infra-red dye labelled sequencing primers were from Li-Cor Biotechnology Division (Lincoln, NB), and Ni-NTA agarose resin was obtained from Qiagen (Mississauga, ON). All aqueous solutions were prepared using water purified with a Millipore BioCell system. DNA sequencing was performed using a kit for Sequenase version 2.0 DNA polymerase (Amersham–Pharmacia Biotech) with a Li-Cor IR2 automated system.

5.3.3 Construction of the expression plasmid

The pQE32-GTG expression plasmid that contains the guinea pig liver TGase gene was constructed using the site overlap extension mutagenesis method [29] starting with the pQE32-TGase plasmid [30]. The pQE32-GTG plasmid differs from pQE32-TGase by the presence of a *Sal*I restriction site at position 799 bp (numbering according to the origin of replication of plasmid pQE32-TGase), the presence of an *Age*I restriction site at position 1728 bp and the absence of a GC-rich 41 bp sequence containing a *Sal*I restriction site 3′ from the stop codon and 5′ from the vector *Pst*I restriction site. To introduce the *Sal*I restriction site at position 799 bp, the guinea pig liver TGase gene was PCR-amplified from plasmid pQE32-TGase using the primer pairs SphIGTGaseF-Met (5′-CACACAGCATGCAGGCAGAGGATCTGATCCTG-3′) with GTG-SalI-799-R (5′-CACGTAGACAG***GTCGAC***TGCGGCGCGAGCA-3′), and GTG-SalI-799-F (5′-TGCTCGCGCCGCA***GTCGAC***CTGTCTACGTG-3′) with pQE32Reverse (5′-GTTCTGAGGTCATTACTGG-3′). These primer pairs introduce a *Sal*I restriction site (bold italic) through two silent mutations (underlined) at codons Ser215 and Arg216 of the guinea pig liver TGase gene. The 2069 bp full length guinea pig liver TGase gene was

generated by a third PCR reaction using the two amplicons and the external primers SphIGTGaseF-Met and pQE32Reverse. This full length PCR product was digested with *Pst*I/*Sph*I, isolated by agarose gel electrophoresis and extracted with a QIAEXII kit (Qiagen). The mutated gene was inserted into pQE32 expression vector (Qiagen) *via* *Sph*I/*Pst*I yielding the plasmid pQE32-TGase(*Sal*I). To introduce the *Age*I restriction site at position 1728 bp, the guinea pig liver TGase gene was PCR-amplified from plasmid pQE32-TGase(*Sal*I) using the primer pairs GTG-AgeI-1728-F (5′-AATGGAGTCCTG*GG**A**CC**G**GT*GTGCAGCACCAAC-3′) with GTG-PstI-2234-R (5′-ACACACCTGCAGTTAGGCGGGGCCGATGAT-3′) and SphIGTGaseF-Met with GTG-AgeI-1728-R (5′-GTTGGTGCTGCAC***AC**C**GGT**CCCAGGACTCCATT-3′). These primer pairs introduce an *Age*I restriction site (bold italic) through two silent mutations (underlined) at codons Gly524 and Pro525 of the guinea pig liver TGase gene while also removing the *Sal*I-containing 41 bp sequence between the stop codon of the TGase gene and the *Pst*I site of plasmid pQE32-TGase(*Sal*I). The 2069 bp full length guinea pig liver TGase gene was generated by a third PCR reaction using the two amplicons and the external primers described above. This PCR product was digested with *Sph*I/*Pst*I, isolated by agarose gel electrophoresis and extracted. The mutated gene was inserted into pQE32 expression vector *via* *Sph*I/*Pst*I yielding the plasmid pQE32-GTG. The entire *Sph*I/*Pst*I fragment that includes the whole coding region was verified by DNA sequencing.

5.3.4 Mutagenesis of Cys336 and Tyr519

The Cys336Ala and Cys336Val mutations of the guinea pig liver TGase gene were introduced by a pair of complementary degenerate oligonucleotides (only the coding strand is shown): GTG-GYT336 (5′-ATCTGGAACTTCCAC**GYT**TGGGTGGAGTCGT-3′). Codon GYT (Y = C or T; bold) encodes both Ala and Val. A third PCR reaction using these two amplicons and the external primers GTG-SalI-799-F and GTG-AgeI-1728-R was performed to obtain the 928 bp guinea pig liver TGase gene fragment located between restriction sites *Sal*I and *Age*I. For mutations Tyr519Phe, Tyr519Leu, Tyr519His and

Tyr519Gln, the primer pair GTG-SalI-799-F with GTG-YWS519 (5′-GAGAACCGGTCCCAGGACTCCATT**SWR**GCTGACGATGCGT-3′) was used. Primer GTG-YWS519 contains the degenerate codon YWS (non-coding strand SWR where S = C or G, W = A or T and R = A or G; bold) that encodes Leu, His, Gln, Phe, Tyr and the amber stop codon. The PCR products were digested by *Sal*I/*Age*I, isolated by agarose gel electrophoresis and extracted. The resulting 928 bp fragments were inserted into pQE32-GTG expression vector *via Sal*I/*Age*I yielding the plasmid pQE32-GTG expressing the mutant TGases. The plasmids were then transformed into *E. coli* XL-1 Blue cells or *E. coli* XL-1 Blue cells containing the plasmid pDnaKJ [31] that encodes the molecular chaperones DnaK and DnaJ. The desired mutants were identified by DNA sequencing and the integrity of the entire coding region was verified. The Cys336 and Tyr519 double mutants were generated by a similar procedure.

5.3.5 Overexpression and purification of wild-type and mutant TGases

Wild-type and mutated recombinant guinea pig liver TGases were overexpressed and purified from *E. coli* according to a protocol developed in our laboratory [30] with the following modifications. After Ni-NTA purification, the eluant was transferred to an Amicon Ultra 15 mL tube (Millipore) with a molecular weight cutoff of 30 kDa for desalting by centrifugation with 25 mM Tris-acetate buffer (pH 7.0) containing 0.5 mM EDTA. The TGase samples were aliquoted at a final concentration of 250 µg/mL, snap-frozen on dry ice and stored at −80 °C. Typical yields were 1.5-10 mg/L of ~85 % pure protein, as estimated from Coomassie Blue staining following SDS-PAGE.

5.3.6 Synthesis of Cbz-Gly-7HC

Cbz-Gly-7HC (Figure 1B) was synthesized based on a previously reported protocol [32]. Namely, 0.2 g (1 mmol) of CbzGly and 0.4 g (2.5 mmol) of 7-hydroxycoumarin were dissolved in 10 mL of ethyl acetate. Then 0.22 mL (0.2 g, 2 mmol) of *N*-methylmorpholine

and 0.8 mL (0.63 g, 5 mmol) of *N,N*-diisopropylcarbodiimide were added with stirring at room temperature. Stirring was continued until the complete disappearance of CbzGly, as followed by thin layer chromatography (ethyl acetate). The reaction mixture was then washed once with 1 M NaOH, three times with 0.1 M NaOH, 3 times with 0.1 M HCl, once with saturated $NaHCO_3$ and once with brine. The organic phase was then dried over $MgSO_4$, filtered and evaporated under reduced pressure. The resulting residue was purified by silica gel chromatography (ethyl acetate) to remove traces of diisopropylurea, giving the desired ester in 70 % yield (0.25 g). ^{1}H NMR (300 MHz, $CDCl_3$) δ 4.30 (2H, d), 5.18 (2H, s), 5.33 (1H, s), 6.44 (1H, d), 7.08 (1H, d), 7.10 (1H, s), 7.37 (5H, m), 7.51 (1H, d), 7.70 (1H, d). ^{13}C NMR (75 MHz, $CDCl_3$) δ 168.5, 160.6, 156.7, 154.9, 152.9, 143.1, 136.3, 129.0, 128.9, 128.6, 128.5, 118.4, 117.2, 116.6, 110.5, 67.7, 43.2. HRMS (FAB) calculated for $C_{19}H_{16}NO_6$ ($[M+H]^+$): 354.0972, found 354.0968.

5.3.7 Kinetic assays

5.3.7.1 Method I: DMPDA assay

The kinetic properties of the mutant TGases were determined using a continuous spectrophotometric assay (Figure 1A) [33]. All kinetic runs were performed in triplicate at six different concentrations of donor substrate. A standard stock buffer solution (182 mM Tris-acetate buffer pH 7.0, 5 mM $CaCl_2$, and 0.9 mM EDTA) was prepared. 0.8 mM of *N,N*-dimethyl-1,4-phenylenediamine (DMPDA) and 0.5-30 mM of *N*-carbobenzyloxyglutaminylglycine (Cbz-Gln-Gly) (Sigma) were added to the stock buffer solution. The reaction solution was preincubated for 3 min at 37 °C prior to initiation of the enzymatic reaction by addition of ~8 µg of purified His_6-TGase. Reaction progress was followed as a linear increase of absorbance measured at 278 nm with a Cary 100 spectrophotometer (Varian) thermostated at 37 °C. For the reaction blanks, water was substituted for TGase enzyme. Slopes were measured over the first <10 % substrate conversion to product and were converted into initial rates by dividing by the extinction

coefficient of the authentic reaction product, Cbz-Glu-γ-[4-(dimethylamino)anilido]Gly ($8940 \pm 55\ M^{-1}cm^{-1}$) [33].

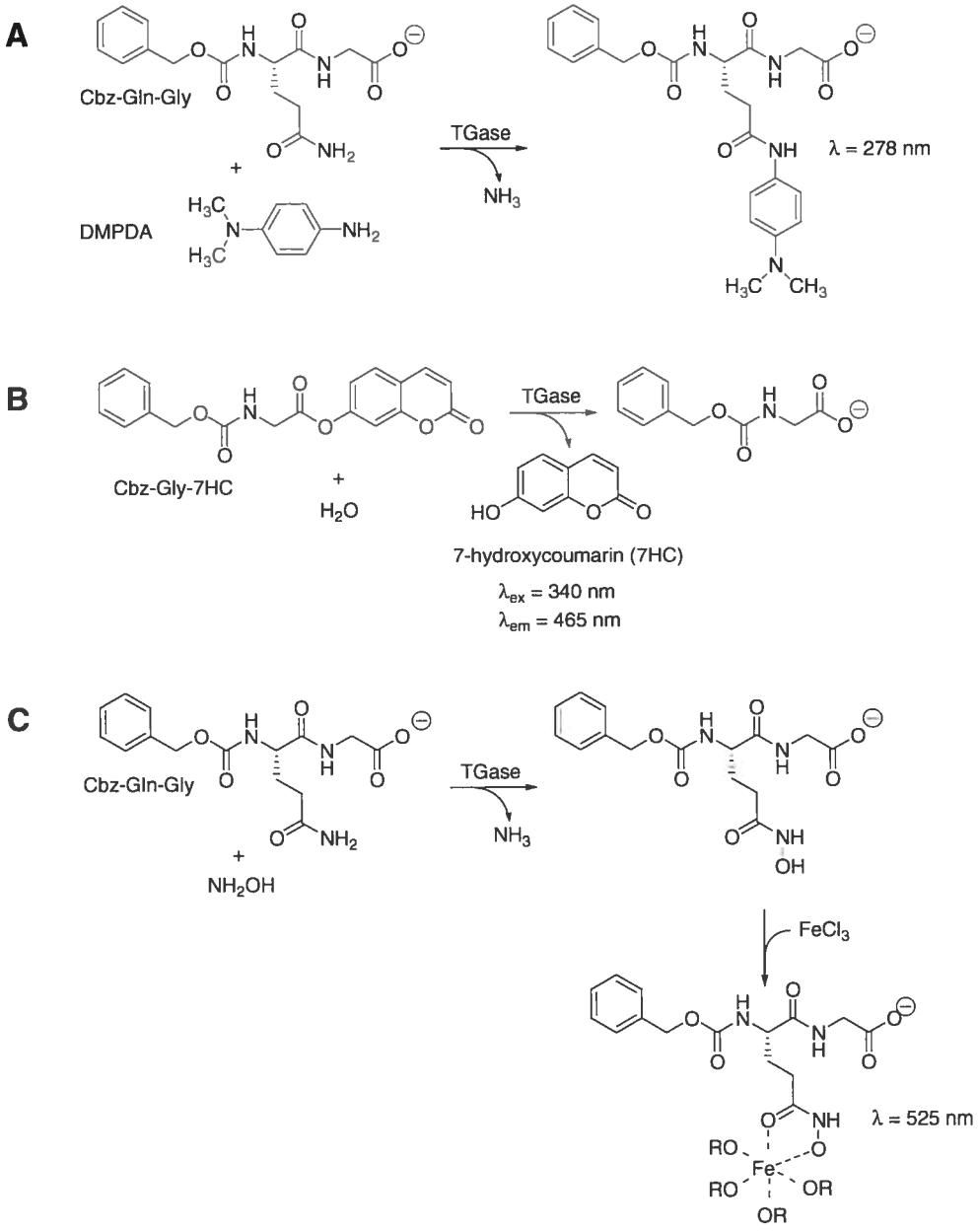


Figure 1. Kinetic assays for tissue TGase used in this work. (A) DMPDA assay. (B) 7-hydroxycoumarin assay. (C) hydroxamate assay.

5.3.7.2 Method II: 7-hydroxycoumarin assay

The decrease of activity of wild-type, Cys336Ala and Cys336Val guinea pig liver TGases as a function of time was measured using the following procedure. Frozen purified enzyme was thawed at room temperature then incubated at 37 °C in its storage buffer. Aliquots of 10 µL (~3 µg) were periodically sampled for immediate measurement of residual activity using a continuous fluorometric assay (Figure 1B) [32] described in the next paragraph. After three hours incubation at 37 °C, dithiothreitol was added to the purified enzyme to a final concentration of 5 mM and a further 10 µL aliquot was sampled.

Residual activity was measured at 25 °C by adding ~3 µg of TGase to each well of a TCT Luminescence 96-well microtiter plate (Thermo Electron) containing a 100 µM solution of the donor substrate Cbz-Gly-7HC in stock buffer solution (100 mM Mops buffer pH 7.0, 5 mM $CaCl_2$, and 0.05 mM EDTA, 5 % DMF). All assays were performed in triplicate. The increase in fluorescence due to the release of 7-hydroxycoumarin was followed against a blank at λ_{ex} 340 nm and λ_{em} 465 nm in a FluoStar Optima microtiter plate reader (BMG Labtech). For the aliquots containing dithiothreitol, a blank containing the same amount of dithiothreitol but with water replacing enzyme was used. Linear slopes of fluorescence versus time were measured over the first <10 % conversion of substrate to product. To convert the slopes into initial rates, a "fluorescence coefficient" was determined, on a daily basis, by measuring the arbitrary fluorescence intensities corresponding to five concentrations of 7-hydroxycoumarin at concentrations ranging from 0.05 to 0.5 µM in 5 % DMF in the stock buffer solution at 25 °C. The value of this "fluorescence coefficient" varied only slightly (<5 %) each day.

5.3.7.3 Method III: Hydroxamate assay

The inactivation of TGase activity as a function of time was also determined using the discontinuous hydroxamate assay (Figure 1C) [34]. Incubation of enzymes at 37 °C and

addition of dithiothreitol after 3 h were as described above. For residual activity assays, a stock solution containing 37.5 mM Cbz-Gln-Gly, 1.25 mM EDTA, 125 mM hydroxylamine and 6.25 mM $CaCl_2$ in 250 mM Tris-acetate buffer (pH 6.0) was prepared. To this solution, ~25 μg of enzyme was added and the reaction was allowed to proceed for 10 minutes after which it was quenched with a ferric chloride reagent solution (1.67 % $FeCl_3 \cdot 6H_2O$, 5 % trichloroacetic acid and 0.83 M HCl). The absorbance was measured at 525 nm on a Biotech Ultrospec 2000 (Pharmacia) table-top spectrophotometer and was converted into initial rates using the extinction coefficient of the hydroxamate-containing product (340 $M^{-1}cm^{-1}$) [21]. For the aliquots containing dithiothreitol, a blank containing the same amount of reducing agent but with water replacing enzyme was used.

5.4 Results and discussion

To test the hypotheses reported in the literature [15, 20] of the putative interactions between the catalytic Cys residue and conserved active-site residues Tyr515 and Cys333 of red sea bream tissue TGase, site-directed mutagenesis of guinea pig liver TGase was envisaged. Guinea pig liver TGase was chosen to test these hypotheses because it is the most widely characterized tissue TGase, allowing correlation with previous results. However, despite numerous studies of this tissue TGase, its structure has not yet been reported. Thus, in order to more confidently guide site-directed mutagenesis of guinea pig liver TGase, a homology model for this enzyme was generated to verify that the positioning of the homologous Tyr and Cys residues (Tyr519 and Cys336, respectively) as well as the overall structure of the enzyme was sufficiently similar to that of red sea bream tissue TGase, from which these functional hypotheses were derived.

5.4.1 Homology modeling

To generate the homology model of guinea pig liver TGase, we used the crystal structures of four TGases as templates: human tissue TGase [14], red sea bream tissue

TGase [15], human epidermal TGase [16] and human factor XIIIa [35]. Guinea pig liver TGase has 83 % residue identity with human tissue TGase, 43 % with red sea bream tissue TGase, 39 % with human epidermal TGase and 36 % with human factor XIIIa. Residue homology is 90 %, 59 %, 56 % and 51 %, respectively, which is reflected in the conserved fold of the four TGases. These structures show sufficiently high sequence and structural similarity to justify their use as templates for a homology model of guinea pig liver TGase.

Sequence alignments between these four TGases and guinea pig liver TGase were performed. Two distinct sequence alignments were generated. The first was derived using the Tcoffee server [23] and the second using ClustalW [24], both with default parameters. As expected from the high homology between the templates sequences, the two resulting alignment files were identical except in the areas defining gaps. The gaps are similar but differ slightly in their extremities and should thus have little effect on homology modeling.

Using the sequence alignments, ten models were generated by Modeller: five using the Tcoffee alignments and five using the ClustalW alignments (Table 1). These models were analyzed using three different criteria: backbone conformation, residue contacts and residue interactions. Backbone conformation was evaluated using Molprobity [26], by verifying torsional angles. A low percentage (≤2.0 %) of residues having torsional angles in the disallowed regions of the Ramachandran plot (Table 1) was obtained, coherent with properly folded models. Furthermore, the percentages obtained for the ten models were in the same range as the values for the template crystal structures. Residue contacts were verified by Whatcheck. No deleterious contacts that were not already present in the template crystal structures were detected, again indicating that the models were well folded. Lastly, analysis of residue interactions was performed using ProsaII [28], an algorithm that measures the interaction energies of residues with the remainder of the protein. All models had regions with unfavourable residue interactions and had to be further refined.

Table 1. Quality evaluation of the guinea pig liver TGase homology models

Crystal structure or model	Ramachandran plot quality prior to refinement (%)			Ramachandran plot quality after refinement (%)			Prosa test after refinement
	Favoured	Allowed	Outlier	Favoured	Allowed	Outlier	
1KV3	91.6	5.8	2.6	N.D.	N.D.	N.D.	N.D.
1G0D	91.8	6.5	1.7	N.D.	N.D.	N.D.	N.D.
1EVU	95.5	4.1	0.4	N.D.	N.D.	N.D.	N.D.
1L9N	96.7	3.2	0.1	N.D.	N.D.	N.D.	N.D.
C1[a]	95.4	3.1	1.5	87.6	9.6	2.8	×[b]
C2	94.3	4.2	1.5	88.4	10.6	1.0	√[b]
C3	95.3	3.1	1.6	88.9	9.8	1.3	×
C4	93.7	4.4	1.9	85.3	12.8	1.9	×
C5	96.0	3.3	0.7	89.6	8.7	1.7	×
T1[a]	94.9	3.8	1.3	89.5	8.9	1.6	√
T2	94.8	3.2	2.0	87.9	11.1	1.0	√
T3	94.8	3.6	1.6	86.9	11.2	1.9	×
T4	94.2	3.9	1.9	87.8	11.2	1.0	×
T5	96.5	2.6	0.9	88.1	9.9	2.0	√

[a] C indicates a model derived from the ClustalW alignment file while T indicates a model derived from the Tcoffee alignment file.

[b] A check mark indicates that residue interactions are favourable according to ProsaII while "×" indicates that they are not.

Model refinement was performed by energy minimization as described under Methods. Following refinement, analysis by Molprobity of the backbone conformation revealed that only four of the ten models (models C2, C3, T2 and T4) had improved their

backbone conformation, as evidenced by a decreased percentage of residues with angles in the disallowed regions of the Ramachandran plot (Table 1). Of these four models, only two (C2 and T2) no longer had regions with unfavourable residue interactions as measured by ProsaII. Analysis of residue contacts by Whatcheck did not reveal any deleterious contacts that were not already present in the templates when refined by the same energy minimization protocol (data not shown).

Models C2 and T2 were selected as the most representative models. Low rmsd values of 1.6 Å for the main chain atoms and 2.2 Å for all heavy atoms were calculated between the two models, demonstrating that these two structures are very similar. Inspection of the active-site residues (Trp241, Cys277, His305, Asp306, Trp332, His335, Cys336, Thr360-Glu363, Tyr519 and Asn520) demonstrates that the active-site structure of the two models is very similar, having an rmsd of 1.3 Å. The conformation of residue Cys336 is different in the two models. In model C2, the χ_1 and χ_2 angles of Cys336 are -176.3° and 61.4°, respectively, while they are -47.8° and -73.1° in model T2. Thus, Cys336 is in a *trans* conformation in model C2 while it is in the most abundant *gauche*$^+$ conformation in model T2. This difference in conformation of the Cys336 side chain modifies the distance between the S_γ atoms of Cys336 and Cys277: it is 5.3 Å for model C2 and 4.5 Å for model T2 (Table 2). These two conformations are allowed according to the χ_1-χ_2 plot for Cys [36] demonstrating that more than one conformation of this residue is allowed in the TGase fold. However, only the *gauche*$^+$ conformation is observed in the four crystal structures used as templates, so model T2 was selected as the best model and analyzed further.

Table 2. Distances between key atoms in human and red sea bream tissue TGases and in homology model T2 of guinea pig liver TGase

Tissue TGase[a]	Cys277 S_γ - Cys336 S_γ[b] (Å)	Cys277 S_γ - Tyr519 O_η[b] (Å)
Human (1KV3)	4.6	2.8
Red sea bream (1G0D)	4.7	2.8
Guinea pig liver (T2)	4.5	2.9

[a] Identified by the PDB code of the crystal structures for human and red sea bream tissue TGases or by the model number of guinea pig liver TGase.

[b] Distances were measured with DS Visualizer v1.5 (Accelrys, San Diego, CA).

A final validation of model T2 was performed by verifying its overall stability by an unconstrained, 300-ps long molecular dynamics simulation at 300 K. If model T2 were not folded properly, it would tend to denature gradually which would be reflected by an increase in the radius of gyration. Figure 2A shows that the total energy of the model decreases rapidly in the first 25 ps and then stabilizes, indicating that the structure is near equilibrium. Figure 2B shows that the radius of gyration of model T2 decreases by ~1 Å in parallel with the energy and then stabilizes during the remainder of the simulation. Together, these results confirm that model T2 of guinea pig liver TGase is stable at room temperature. Model T2 was thus validated and used for comparison with the crystal structure of red sea bream tissue TGase.

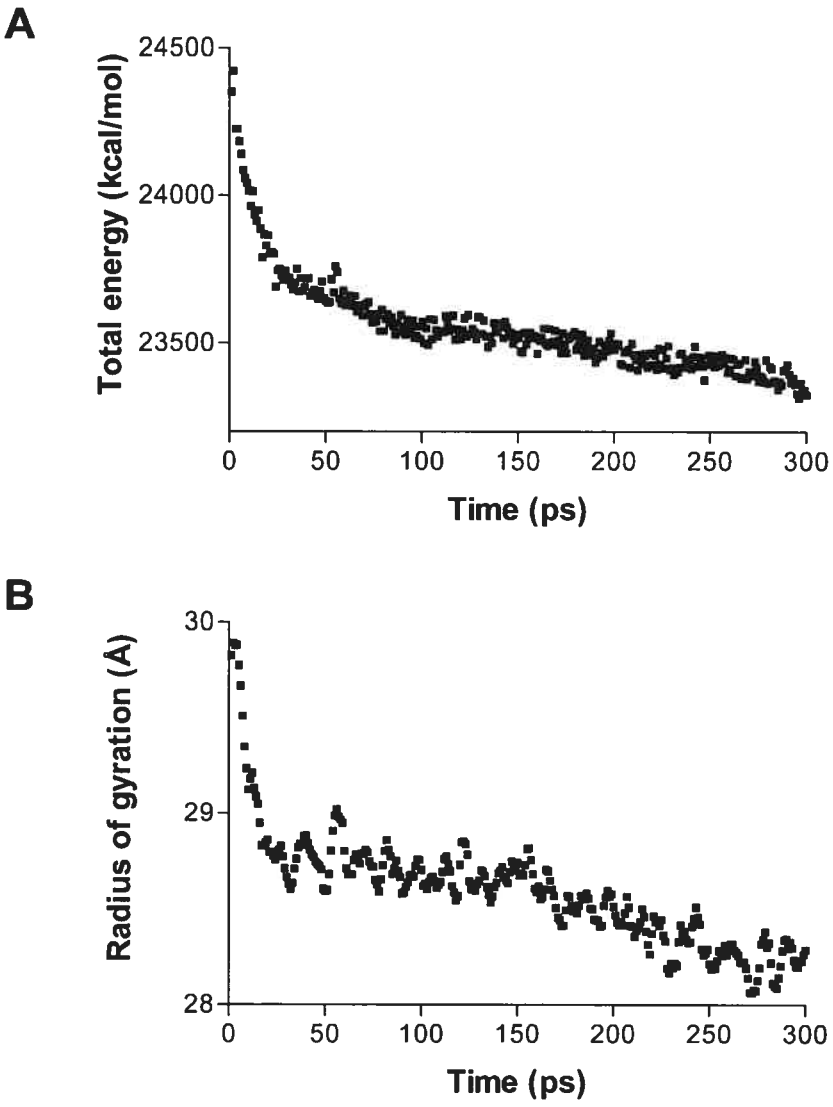


Figure 2. Stability of guinea pig liver TGase homology model T2 during an unconstrained 300 ps molecular dynamics simulation at 300 K. Stability was evaluated by measuring the total energy (A) and the radius of gyration (B).

Analysis of model T2 reveals that the overall structure of guinea pig liver TGase contains the four domains common to other TGases (Figure 3), namely the N-terminal β-sandwich domain, the catalytic core containing the conserved catalytic triad, the barrel 1

domain and the C-terminal barrel 2 domain. Model T2 is very similar to human tissue TGase, having an rmsd of 1.4 Å for the backbone and 3.0 Å for the heavy atoms of residues Cys277, Cys336 and Tyr519. Model T2 is also similar to red sea bream tissue TGase, having an rmsd of 4.0 Å for the backbone and 3.0 Å for the heavy atoms of residues Cys277, Cys336 and Tyr519. These data are consistent with guinea pig liver TGase having a higher sequence homology with human tissue TGase than with red sea bream tissue TGase. The structural similarity between homology model T2 and the template crystal structures is not surprising, as the four initial templates are themselves structurally similar and share high sequence homology.

The high similarity between the active-sites of model T2, human tissue TGase and red sea bream tissue TGase is further demonstrated by the fact that the active-site region is negatively charged for all structures (Figure 3) and that the overall charge distribution is similar. Furthermore, the distances between the S_γ atom of the catalytic Cys277 residue and the S_γ of Cys336 or the O_η of Tyr519 are similar between the tissue TGases (Table 2), indicating a high similarity between the active-site structures. These observations validate the use of guinea pig liver TGase, the catalytic mechanism of which has been the most extensively characterized, for testing the putative interactions that have been postulated to occur between the catalytic Cys and conserved active-site Tyr and Cys residues.

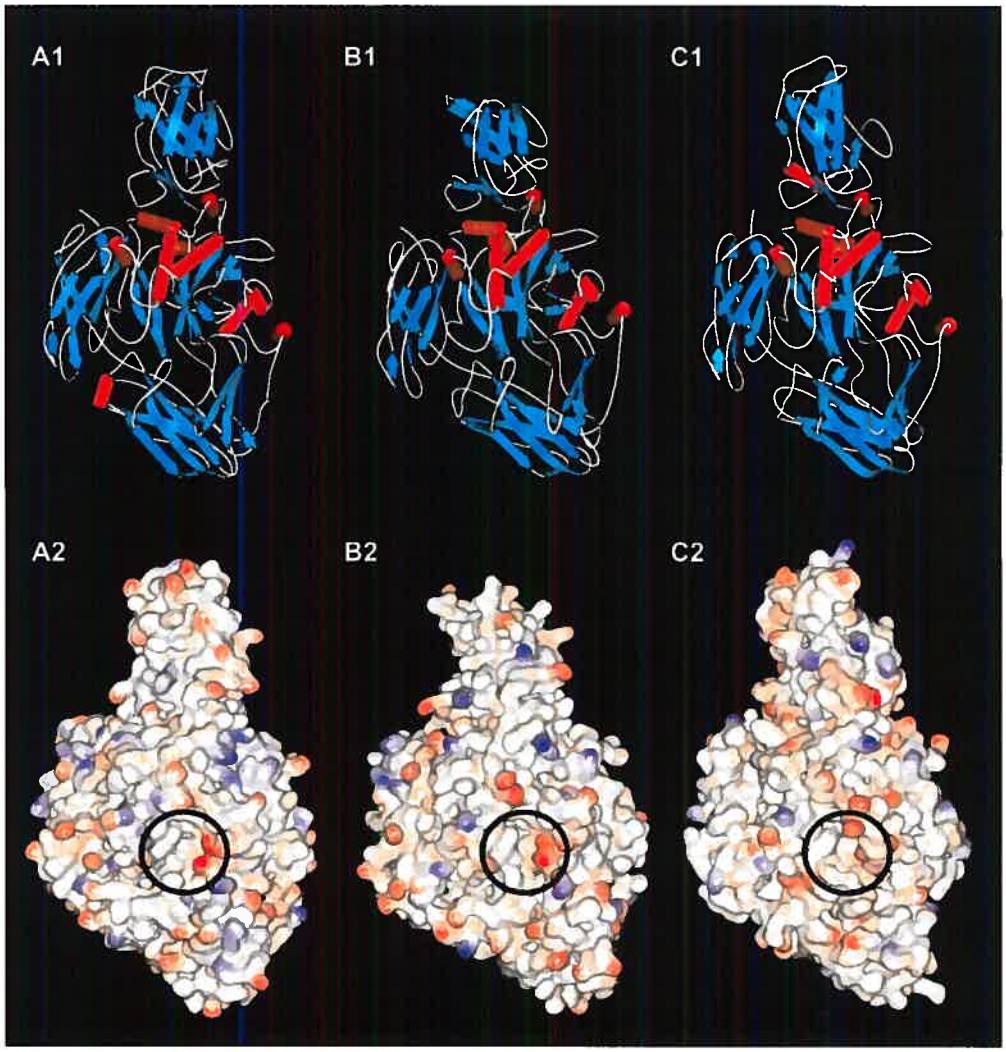


Figure 3. Structures of various tissue TGases. (A) Model T2 of guinea pig liver TGase. (B) Crystal structure of human tissue TGase (PDB code 1KV3). (C) Crystal structure of red sea bream tissue TGase (PDB code 1G0D). A1–C1: schematic representation of the secondary structure elements. A2–C2: surface representation (blue represents positive charges, red represents negative charges). The active-site area is circled.

We have thus generated the first homology model of guinea pig liver TGase for which no crystal structure has been reported. This enzyme is the most widely-studied TGase and has long served as a model system for the study of tissue TGases. Our

homology model will find general application in the study of structure-function relationships and in the design of new substrates and inhibitors.

5.4.2 Construction of plasmid pQE32-GTG

Expression plasmid pQE32-GTG was constructed to facilitate our mutagenesis experiments on the ~2000-bp long guinea pig liver TGase gene. The *Sal*I and *Age*I restriction sites were introduced at positions 799 bp and 1728 bp of plasmid pQE32-TGase through silent mutations of the guinea pig liver TGase gene to allow PCR mutagenesis of a 928-bp DNA fragment instead of the entire ~2000-bp gene. This precaution was taken to minimize the occurrence of non-specific mutations during the subsequent mutagenic site overlap extension PCR. In addition, the 41-bp sequence between the stop codon of the guinea pig liver TGase gene and the *Pst*I site of plasmid pQE32-TGase was removed to eliminate an undesired *Sal*I restriction site and to facilitate DNA sequencing from the 3′ end of the gene which had been hindered due to its high GC content (78 %, including a 12-bp G repeat). These mutations have no effect on the amino acid sequence of the wild-type guinea pig liver TGase, nor on its specific activity when expressed in *E. coli* (data not shown).

5.4.3 Role of guinea pig liver TGase residue Cys336

It has been reported that guinea pig liver TGase loses activity steadily during storage and that much of the lost activity can be restored by addition of 1-5 mM of dithiothreitol [21, 22]. This suggests that oxidation is the cause of inactivation. For this reason, we store TGase samples at -80 °C in small aliquots that are thawed and discarded after a 1-day experiment. Recently, it has been proposed that the S_γ atom of the catalytic Cys residue of red sea bream tissue TGase (Cys272) can form a disulphide bond with the S_γ atom of a second conserved Cys residue (Cys333) located 4.7 Å away according to the crystal structure [15]. Disulphide bond formation would inactivate the enzyme. This hypothesis is consistent with the time-dependent loss of activity of guinea pig liver TGase,

also a tissue TGase. To test the hypothesis of oxidative enzyme inactivation *via* disulphide bridge formation, we mutated the second conserved Cys in guinea pig liver TGase (Cys336, corresponding to red sea bream tissue TGase Cys333). The guinea pig liver enzyme was mutated as it is the most extensively characterized tissue TGase, allowing a correlation with previous mechanistic work. Cys336 was mutated to Ala or Val; no undesired secondary mutations were observed. We determined the catalytic parameters of the Cys336 mutants using the DMPDA assay (Table 3) because it uses a glutamine-containing peptide as a donor substrate, closely resembling the natural substrates of the enzyme. The catalytic efficiency of both mutants is approximately 50 % that of the wild-type enzyme. For mutant Cys336Ala, an increase in K_M accounts for most of the decreased catalytic efficiency, while a decrease in k_{cat} accounts for most of the decreased activity in the case of mutant Cys336Val. These results confirm that residue Cys336 does not play a direct role in catalysis nor does the mutation of this active-site residue have a large effect on activity.

Table 3. Kinetic constants for Cbz-Gln-Gly of wild-type and mutant guinea pig liver TGases determined using the DMPDA method

TGase	K_M (mM)	k_{cat} (min^{-1})	k_{cat}/K_M ($min^{-1}mM^{-1}$)	k_{cat}/K_M mutant / k_{cat}/K_M WT
Wild-type	7 ± 1	116 ± 7	17	1.0
Cys336Ala	17 ± 2	110 ± 7	6	0.4
Cys336Val	5 ± 1	47 ± 5	9	0.5
Tyr519Phe	8 ± 1	85 ± 4	11	0.6
Tyr519His	4 ± 2	19 ± 2	5	0.3
Tyr519Leu	6 ± 2	67 ± 10	11	0.6
Tyr519Gln	7 ± 2	48 ± 4	7	0.4
Cys336Val + Tyr519His	5 ± 1	44 ± 5	9	0.5
Cys336Val + Tyr519Leu	8 ± 2	43 ± 5	5	0.3
Cys336Val + Tyr519Gln	4.3 ± 0.6	46 ± 2	11	0.6

In order to assess the potential role of Cys336 in the putative time-dependent oxidative loss of activity, the catalytic activity of wild-type, Cys336Ala and Cys336Val guinea pig liver TGases was determined at various times during a 3 h incubation at 37 °C (Figure 4). We monitored activity using the 7-hydroxycoumarin assay [32] because it is continuous, simple and sensitive enough to be performed in a microtiter plate format. When we first performed the experiment, the enzymes were incubated at 37 °C in the presence of 5 mM $CaCl_2$ since Ca^{2+} is required to activate the enzyme. However, TGase precipitated when incubated under these conditions (data not shown). Folk and Cole observed previously that incubation of guinea pig liver TGase at 25 °C in 5 mM $CaCl_2$ at pH 6 for 5 min resulted in approximately 20 % loss in activity [22]. Our observations lead us to propose that a conformational change upon Ca^{2+} binding [37-39] coupled to the chaotropic calcium reduces TGase solubility. When the incubation was repeated without calcium, a time-dependent loss of activity was observed, in the absence of any apparent protein aggregation. After 3 h incubation, wild-type TGase lost less activity than the mutants. It retained 46 % activity while Cys336Ala and Cys336Val retained only 14 % and 8 % activity, respectively (Figure 4). This more pronounced loss of activity by the two mutants was unexpected: if Cys336 participated in oxidative disulphide bond formation with the catalytic Cys, then the Cys336Ala/Val mutants should have remained active.

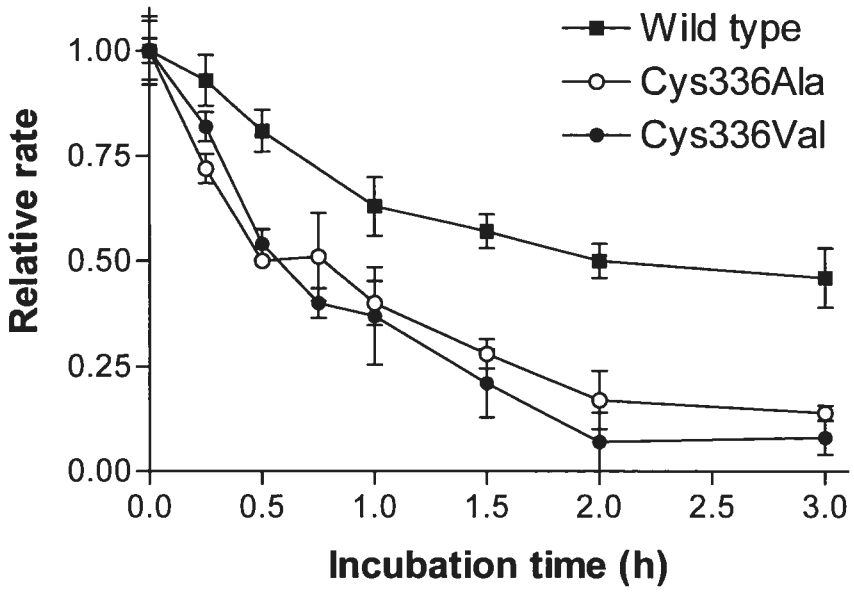


Figure 4. Activity loss of wild-type and Cys336 mutants of guinea pig liver TGase upon incubation at 37 °C. The assay used to measure activity was the 7-hydroxycoumarin assay.

Following incubation at 37 °C for 3 h, 5 mM dithiothreitol was added to the enzyme aliquots. Dithiothreitol was expected to reduce the putative Cys277-Cys336 disulphide bond in the inactivated wild-type TGase. However, dithiothreitol did not allow recovery of activity for the wild-type TGase (data not shown). While the 7-hydroxycoumarin assay is rapid and sensitive, it features an aromatic ester as a donor substrate for TGase. This ester is also sensitive to nucleophilic attack by dithiothreitol. To eliminate the possibility that the experimental controls were flawed, the experiment was repeated using the hydroxamate assay which is not sensitive to dithiothreitol. Using this assay, we observed that after 3 h incubation, wild-type TGase retained 53 % activity while Cys336Ala and Cys336Val retained 25 % and 12 % activity, respectively, consistent with our previous results. The confirmation that mutants Cys336Ala and Cys336Val are inactivated more rapidly than the wild-type invalidates the hypothesis of oxidation of the catalytic Cys residue through the

formation of a disulphide bridge with Cys336. Furthermore, according to our homology model, no other Cys residue is located within disulphide-forming distance of the catalytic Cys residue. Again, addition of 5 mM dithiothreitol did not result in the recovery of activity of the wild-type TGase. These results demonstrate that inactivation of TGase over time does not result from formation of a Cys277-Cys336 disulphide bond. However, oxidation of the catalytic Cys residue to a species that cannot be reduced by dithiothreitol is still possible. Not unexpectedly, no reactivation of the Cys336Ala/Val mutants was achieved by addition of dithiothreitol. The more rapid inactivation and greater extent of inactivation of the two mutants may result from less stable folding.

5.4.4 Role of guinea pig liver TGase residue Tyr519

Molecular modeling performed on the crystal structure of red sea bream tissue TGase has suggested that residue Tyr515 forms a critical H-bond with the donor substrate Cbz-Gln-Gly during the acylation step of the TGase mechanism [20]. This H-bond would orient the substrate γ-carboxamide group for nucleophilic attack by the catalytic thiol (Cys272) and increase electrophilicity of the γ-carboxamide, making it more reactive. After nucleophilic attack, the resulting oxyanion would also be stabilized by the H-bond formed with the hydroxyl group of the conserved Tyr515, stabilizing the tetrahedral intermediate formed during the acylation step. To test this hypothesis, homologous residue Tyr519 of guinea pig liver TGase was mutated to amino acids that cannot form a H-bond with the substrate (Phe or Leu), as well as to amino acids that have the potential to form a H-bond with the substrate (His or Gln). According to our hypothesis [20], if loss of the H-bond reduces the rate of the acylation step to the point where it becomes rate-limiting, the Leu and Phe mutants would lose activity while the Gln and His mutants may retain activity.

The DMPDA kinetic assay was used to measure the effect of these mutations on the catalytic activity. This assay was chosen because it is continuous, direct and, most importantly, uses the modelled Cbz-Gln-Gly donor substrate that would require the

proposed H-bond with Tyr519 to be activated. This is not the case for direct continuous assays that use aromatic esters as donor substrates such as the 7-hydroxycoumarin assay [32] or the Cbz-L-Glu(γ-*p*-nitrophenyl ester)Gly assay [8]. All single mutants showed a two- to four-fold decrease in catalytic efficiency relative to wild-type, mostly as a result of decreased k_{cat} (Table 3). Thus, contrary to our hypothesis regarding residue Tyr519, amino acids with different potentials for H-bond formation yielded similar decreases in activity. Clearly, Tyr519 is not critical for catalytic activity, nor does it appear that H-bonding by Tyr519 plays a large role in stabilization of the transition state of the rate-limiting deacylation step. Furthermore, contrary to what has been proposed in the literature [15], mutation of this Tyr to an amino acid that cannot form H-bonds does not lead to immediate inactivation of tissue TGase *via* disulphide bridging of the catalytic Cys.

We [8] and others [9] have observed the acceleration of the steady-state rate of transamidation mediated by TGase, using a series of increasingly nucleophilic acceptor substrates. The fact that the steady-state rate increases without attaining a plateau of activity is consistent with the acylation step being faster than deacylation, even with the most nucleophilic acceptor substrates chosen [8, 9]. We postulate that Tyr519 may play an important role in activation of the donor substrate during the acylation step, but the loss of activity caused by the Tyr519 mutations is not sufficient to render the acylation step significantly slower than the deacylation step. The observed modest decreases in the steady-state rate constants are more likely due to the slight decrease in rate of deacylation, through the disruption of transition state stabilisation by H-bonds or other dipole or hydrophobic interactions. The activity assays used in this study are not amenable to measuring the rate constants of the separate acylation and deacylation steps, but we are currently undertaking rapid-mix quench experiments to quantitatively analyze the "burst" phase to measure rate constants for the acylation step of these mutants.

Mutation of Tyr519 in combination with Cys336 yields results similar to mutation of either residue alone, indicating that mutation of two residues in very close proximity to

the catalytic Cys residue does not inactivate the enzyme. No co-operative effect is observed that is different from the results of the single mutations. Our homology model and the crystal structures available for other TGases suggest that these residues are in close proximity to the catalytic Cys residue, so it is surprising that they can be combinatorially replaced and yield mutants with near-native activity. Thus, the active-site of guinea pig liver TGase appears to be permissive to mutations of non-catalytic residues.

5.5 Conclusion

In summary, we have generated the first model of guinea pig liver TGase, the most widely studied tissue TGase. The quality of this model was validated using four different criteria: backbone conformation, residue contacts, residue interactions and overall stability. All values for these criteria fall within the limits of reliable structures. The model was similar to the crystal structures of other tissue TGases, validating the use of guinea pig liver TGase for mutagenesis experiments based on structure/function considerations. Mutation of active-site residues Cys336 and Tyr519, alone or in combination, yielded mutant enzymes that had only a two- to four-fold decrease in catalytic efficiency relative to wild-type, suggesting that the active-site of tissue TGase is tolerant to mutations. Furthermore, we demonstrated that residue Cys336 does not form an inactivating disulphide bridge with the catalytic Cys residue as had previously been suggested.

5.6 Acknowledgements

R.A.C. was recipient of a Natural Sciences and Engineering Research Council of Canada graduate scholarship. This research was funded by a grant from the Fonds Québécois de la recherche sur la nature et les technologies (jointly held by J.N.P. and J.W.K.).

5.7 References

1 Lorand, L. and Graham, R. M. (2003) Transglutaminases: crosslinking enzymes with pleiotropic functions. Nature Reviews Molecular Cell Biology **4**, 140-156

2 Griffin, M., Casadio, R. and Bergamini, C. M. (2002) Transglutaminases: nature's biological glues. Biochemical Journal **368**, 377-396

3 Beninati, S. and Piacentini, M. (2004) The transglutaminase family: an overview: minireview article. Amino Acids **26**, 367-372

4 Fesus, L. and Piacentini, M. (2002) Transglutaminase 2: an enigmatic enzyme with diverse functions. Trends in Biochemical Sciences **27**, 534-539

5 Mehta, K. (2005) Mammalian transglutaminases: a family portrait. Progress in Experimental Tumor Research **38**, 1-18

6 Connellan, J. M. and Folk, J. E. (1969) Mechanism of the inactivation of guinea pig liver transglutaminase by 5,5'-dithiobis-(2-nitrobenzoic acid). Journal of Biological Chemistry **244**, 3173-3181

7 Aeschlimann, D. and Paulsson, M. (1994) Transglutaminases: protein cross-linking enzymes in tissues and body fluids. Thromb. Haemost. **71**, 402-415

8 Leblanc, A., Gravel, C., Labelle, J. and Keillor, J. W. (2001) Kinetic studies of guinea pig liver transglutaminase reveal a general-base-catalyzed deacylation mechanism. Biochemistry **40**, 8335-8342

9 Case, A. and Stein, R. L. (2003) Kinetic analysis of the action of tissue transglutaminase on peptide and protein substrates. Biochemistry **42**, 9466-9481

10 Fox, B. A., Yee, V. C., Pedersen, L. C., Le Trong, I., Bishop, P. D., Stenkamp, R. E. and Teller, D. C. (1999) Identification of the calcium binding site and a novel ytterbium site in blood coagulation factor XIII by x-ray crystallography. Journal of Biological Chemistry **274**, 4917-4923

11 Pedersen, L. C., Yee, V. C., Bishop, P. D., Le Trong, I., Teller, D. C. and Stenkamp, R. E. (1994) Transglutaminase factor XIII uses proteinase-like catalytic triad to crosslink macromolecules. Protein Science **3**, 1131-1135

12 Yee, V. C., Pedersen, L. C., Le Trong, I., Bishop, P. D., Stenkamp, R. E. and Teller, D. C. (1994) Three-dimensional structure of a transglutaminase: human blood coagulation factor XIII. Proceedings of the National Academy of Sciences of the United States of America **91**, 7296-7300

13 Weiss, M. S., Metzner, H. J. and Hilgenfeld, R. (1998) Two non-proline cis peptide bonds may be important for factor XIII function. FEBS Letters **423**, 291-296

14 Liu, S., Cerione, R. A. and Clardy, J. (2002) Structural basis for the guanine nucleotide-binding activity of tissue transglutaminase and its regulation of transamidation activity. Proceedings of the National Academy of Sciences of the United States of America **99**, 2743-2747

15 Noguchi, K., Ishikawa, K., Yokoyama, K., Ohtsuka, T., Nio, N. and Suzuki, E. (2001) Crystal structure of red sea bream transglutaminase. Journal of Biological Chemistry **276**, 12055-12059

16 Ahvazi, B., Kim, H. C., Kee, S. H., Nemes, Z. and Steinert, P. M. (2002) Three-dimensional structure of the human transglutaminase 3 enzyme: binding of calcium ions changes structure for activation. EMBO Journal **21**, 2055-2067

17 Ahvazi, B., Boeshans, K. M. and Steinert, P. M. (2004) Crystal structure of transglutaminase 3 in complex with GMP: structural basis for nucleotide specificity. Journal of Biological Chemistry **279**, 26716-26725

18 Ahvazi, B., Boeshans, K. M., Idler, W., Baxa, U. and Steinert, P. M. (2003) Roles of calcium ions in the activation and activity of the transglutaminase 3 enzyme. Journal of Biological Chemistry **278**, 23834-23841

19 Ahvazi, B., Boeshans, K. M., Idler, W., Baxa, U., Steinert, P. M. and Rastinejad, F. (2004) Structural basis for the coordinated regulation of transglutaminase 3 by

guanine nucleotides and calcium/magnesium. Journal of Biological Chemistry **279**, 7180-7192

20 Chica, R. A., Gagnon, P., Keillor, J. W. and Pelletier, J. N. (2004) Tissue transglutaminase acylation: Proposed role of conserved active-site Tyr and Trp residues revealed by molecular modeling of peptide substrate binding. Protein Science **13**, 979-991

21 Folk, J. E. and Chung, S. I. (1985) Transglutaminases. Methods in Enzymology **113**, 358-375

22 Folk, J. E. and Cole, P. W. (1966) Transglutaminase: mechanistic features of the active-site as determined by kinetic and inhibitor studies. Biochimica & Biophysica Acta **122**, 244-264

23 Notredame, C., Higgins, D. G. and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. Journal of Molecular Biology **302**, 205-217

24 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research **22**, 4673-4680

25 Sali, A. and Blundell, T. L. (1993) Comparative Protein Modeling by Satisfaction of Spatial Restraints. Journal of Molecular Biology **234**, 779-815

26 Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. and Richardson, D. C. (2003) Structure validation by Calpha geometry: phi,psi and C beta deviation. Proteins **50**, 437-450

27 Hooft, R. W. W., Vriend, G., Sander, C. and Abola, E. E. (1996) Errors in Protein Structures. Nature **381**, 272

28 Sippl, M. J. (1993) Recognition of Errors in 3-Dimensional Structures of Proteins. Proteins **17**, 355-362

29 Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene **77**, 51-59

30 Gillet, S. M., Chica, R. A., Keillor, J. W. and Pelletier, J. N. (2004) Expression and rapid purification of highly active hexahistidine-tagged guinea pig liver transglutaminase. Protein Expression & Purification **33**, 256-264

31 Yokoyama, K., Kikuchi, Y. and Yasueda, H. (1998) Overproduction of DnaJ in *Escherichia coli* improves *in vivo* solubility of the recombinant fish-derived transglutaminase. Bioscience, Biotechnology & Biochemistry **62**, 1205-1210

32 Gillet, S. M., Pelletier, J. N. and Keillor, J. W. (2005) A direct fluorometric assay for tissue transglutaminase. Analytical Biochemistry **347**, 221-226

33 de Macedo, P., Marrano, C. and Keillor, J. W. (2000) A direct continuous spectrophotometric assay for transglutaminase activity. Analytical Biochemistry **285**, 16-20

34 Folk, J. E. and Cole, P. W. (1966) Identification of a functional cysteine essential for the activity of guinea pig liver transglutaminase. Journal of Biological Chemistry **241**, 3238-3240

35 Garzon, R. J., Pratt, K. P., Bishop, P. D., Le Trong, I., Stenkamp, R. E. and Teller, D. C. (2000) Tryptophan 279 is Essential for the Transglutaminase Activity of Coagulation Factor XIII: Functional and Structural Characterization. To be published

36 Janin, J. and Wodak, S. (1978) Conformation of amino acid side-chains in proteins. Journal of Molecular Biology **125**, 357-386

37 Casadio, R., Polverini, E., Mariani, P., Spinozzi, F., Carsughi, F., Fontana, A., Polverino de Laureto, P., Matteucci, G. and Bergamini, C. M. (1999) The structural basis for the regulation of tissue transglutaminase by calcium ions. European Journal of Biochemistry **262**, 672-679

38 Di Venere, A., Rossi, A., De Matteis, F., Rosato, N., Agro, A. F. and Mei, G. (2000) Opposite effects of Ca(2+) and GTP binding on tissue transglutaminase tertiary structure. Journal of Biological Chemistry **275**, 3915-3921

39 Mariani, P., Carsughi, F., Spinozzi, F., Romanzetti, S., Meier, G., Casadio, R. and Bergamini, C. M. (2000) Ligand-induced conformational changes in tissue transglutaminase: Monte Carlo analysis of small-angle scattering data. Biophysical Journal **78**, 3240-3251

CHAPITRE 6

Évolution dirigée de la spécificité de la TGase de foie de cobaye par une approche semi-aléatoire

6.0 Préface

À la suite de l'obtention d'information structurelle et fonctionnelle et du développement d'une méthode d'expression et de purification de la TGase de foie de cobaye, il nous fallait développer une méthode de criblage pour détecter les mutants ayant une spécificité modifiée, afin de pouvoir faire l'évolution dirigée de l'enzyme. Dans ce chapitre, nous décrivons la création de cinq banques de mutants et leur criblage avec de nouveaux composés fluorogènes, durant le processus d'évolution dirigée selon une approche semi-aléatoire de la TGase de foie de cobaye.

Pour le manuscrit « Expansion of the peptide synthase specificity of guinea pig liver transglutaminase by semi rational mutagenesis » présenté au chapitre 6, j'ai effectué la majorité des expériences d'expression et de purification des TGases, de cinétique enzymatique, de construction des banques de mutants et de criblage, en plus de rédiger le manuscrit. Jessica Laroche et Philippe Lesage m'ont assisté dans les expériences d'expression et de purification des TGases, de cinétique enzymatique, de construction des banques de mutants et de criblage en tant que stagiaires sous ma supervision. Le Dr Steve Gillet a développé le protocole de criblage en plaques à 96 puits en plus de synthétiser les composés Cbz-Gly-7HC, Cbz-Ala-7HC et Cbz-Phe-7HC. Le Pr Joelle Pelletier a effectué les calculs de représentativité des banques de mutants.

Article 5.

Expansion of the peptide synthase specificity of guinea pig liver transglutaminase by semi-random mutagenesis

Roberto A. Chica, Steve M.F.G. Gillet, Jessica Laroche, Philippe Lesage, Jeffrey W. Keillor and Joelle N. Pelletier

Département de chimie, Université de Montréal, CP 6128, Succursale Centre-Ville, Montréal, Québec, H3C 3J7, Canada

Manuscrit prêt à être soumis

6.1 Abstract

Transglutaminases (TGases) catalyze the cross-linking of proteins and peptides by the formation of γ-glutamyl-ε-lysyl bonds between an acyl-donor and an acyl-acceptor substrate. *In vivo*, the acyl-donor substrate is a glutamine-containing protein while the acyl-acceptor is a lysine-containing protein. The acyl-acceptor substrate can be replaced *in vitro* by primary amines. Recently, we demonstrated the use of a novel class of acyl-donor substrates of guinea pig liver TGase that contain 7-hydroxycoumarin esters. In this study, we explored the substrate specificity of guinea pig liver TGase by assessing its reactivity toward a variety of potential acyl-donor substrates having a 7-hydroxycoumarin ester on the α-carboxyl group of various amino acids. We also assessed its reactivity toward a variety of amino acid derivatives as acyl-acceptor substrates. We observed that the wild-type enzyme can react directly at the α-carboxyl group of *N*-carbobenzyloxyglycine coumarin-7-yl ester, demonstrating for the first time that guinea pig liver TGase has a peptide synthase activity as it can catalyze the formation of peptide bonds between Gly and a series of small polar and hydrophobic amino acid derivatives. To further broaden its specificity, we performed semi-random mutagenesis by performing saturation and often combinatorial mutagenesis at conserved active-site residues (Trp241; His305 and Asp306; Trp332; Thr360, Gln362 and Glu363; Tyr519 and Asn520). We developed a microtiter plate-based screening assay that allowed identification of mutants with altered specificity. Mutant Trp332Phe recognized L-Phe and L-Tyr derivatives as acyl-acceptor substrates while the wild-type enzyme did not, demonstrating that semi-random mutagenesis of the active-site rapidly yielded changes in acyl-acceptor substrate specificity.

Keywords: tissue transglutaminase, TGase, semi-random mutagenesis, peptide synthase, combinatorial mutagenesis, enzyme kinetics

6.2 Introduction

Transglutaminases (TGases, EC 2.3.2.13) catalyze the Ca^{2+}-dependent cross-linking of peptides and proteins *via* the formation of γ-glutamyl-ε-lysyl isopeptide bonds [1-3]. The catalytic reaction follows a modified ping-pong mechanism in which a glutamine-containing protein or peptide, the acyl-donor substrate, reacts with the catalytic cysteine residue to form a thioester bond, thus generating a covalent acyl-enzyme intermediate. The acyl-enzyme intermediate then reacts with a second substrate, the acyl-acceptor, to yield the isopeptide-containing product and free enzyme in a transamidation reaction. In the absence of an acyl-acceptor, the acyl-enzyme intermediate is hydrolyzed, transforming the acyl-donor glutamine residue into glutamate and regenerating the free enzyme [4]. Catalysis is carried out by a Cys-His-Asp catalytic triad similar to that of cysteine proteases. TGases are divided into nine classes [5] of which the ubiquitous tissue TGase found in all vertebrates [6] is the most widely distributed in eukaryotic organisms. Tissue TGases, such as guinea pig liver TGase, are intracellular and monomeric enzymes whose transamidation activity is inhibited by GTP.

Tissue TGases exhibit a broad specificity towards the acyl-acceptor as almost any primary amine can serve as a substrate [7] although in nature it is generally a lysine-containing protein or peptide. Many primary amines, such as glycinamide [8, 9], and anilines, such as *N,N*-dimethyl-1,4-phenylenediamine [10], can also function as acyl-acceptor substrates. However, amines containing free carboxylic acid groups, such as free amino acids, and amines with adjacent bulky substituents, such as L-tyrosinamide, do not act as substrates [9].

Tissue TGase displays narrower specificity for its acyl-donor substrates. The side chain of protein-bound L-Gln residue is the native substrate while the side chains of similar amino acids such as L-Asn and D-Gln are not acyl-donor substrates [8]. Synthetic peptides containing L-Gln are also acyl-donor substrates of TGases. For this reason, the dipeptide

N-carbobenzyloxyglutaminylglycine (Cbz-Gln-Gly, Figure 1A) has become a widely used acyl-donor substrate of tissue TGase. Amides are not the only acyl-donor substrates of TGases: γ-glutamyl aromatic ester derivatives of L-Glu, such as *N*-carbobenzyloxy-L-glutamyl(γ-*p*-nitrophenyl ester)glycine (Cbz-Glu(γPNP)-Gly, Figure 1B), have also been shown to be acyl-donor substrates of tissue TGase and are used to measure the enzyme's activity [11]. However, secondary amides derivatives of L-Gln, such as *N*-δ-methyl-L-glutamine or anilides, are not substrates of tissue TGase [12]: the γ-carboxamide group of L-Gln is the only known amide that is an acyl-donor substrate of tissue TGase.

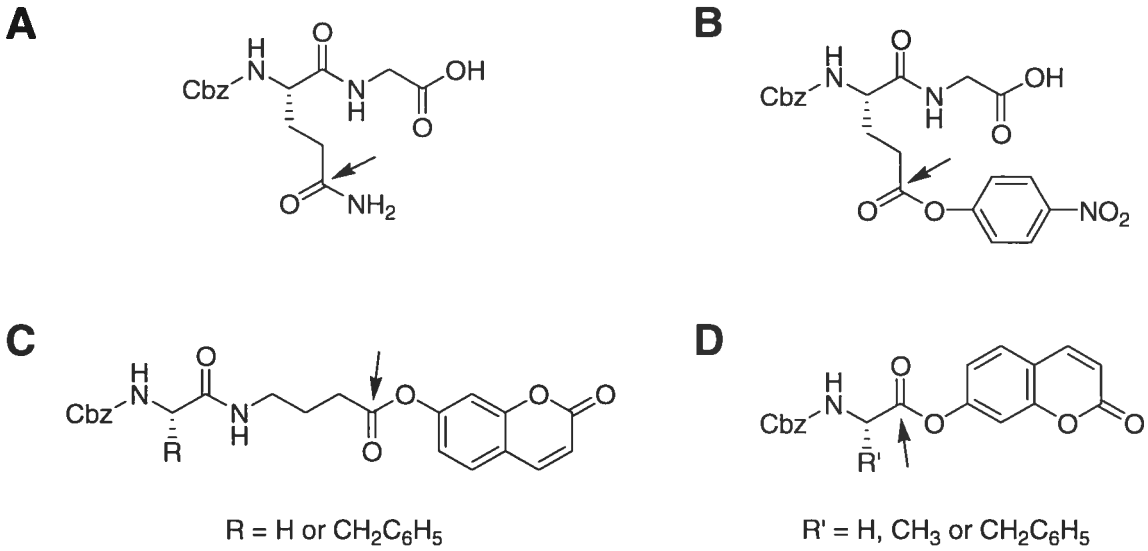


Figure 1. Acyl-donor substrates of guinea pig liver TGase. Cbz is the carbobenzyloxy protecting group. The reactive carbonyl group of these compounds is indicated by an arrow. (A) Cbz-Gln-Gly; (B) Cbz-Glu(γPNP)-Gly; (C) Cbz-Gly-GABA-7HC and Cbz-Phe-GABA-7HC; (D) Cbz-Gly-7HC, Cbz-Ala-7HC and Cbz-Phe-7HC.

Recently, we demonstrated the use of a novel class of acyl-donor substrates of guinea pig liver TGase that are neither L-Gln or L-Glu derivatives [13]. These novel

substrates, 4-(*N*-carbobenzyloxyglycylamino)-butyric acid coumarin-7-yl ester and 4-(*N*-carbobenzyloxyphenylalanylamino)-butyric acid coumarin-7-yl ester (Cbz-Gly-GABA-7HC and Cbz-Phe-GABA-7HC, respectively, Figure 1C), have scaffolds that resemble those of recently published TGase irreversible inhibitors [14, 15]. Upon reaction with TGase, 7-hydroxycoumarin is released, resulting in an increase in fluorescence that makes these compounds useful for measuring the rates of the hydrolysis or transamidation reactions catalyzed by guinea pig liver TGase. The scaffolds of these substrates differ from the typical L-Glu aromatic ester acyl-donor substrates of tissue TGases in that the reactive ester function is located on the main chain instead of the side chain, thus giving rise to products that do not contain a γ-glutamyl-ε-lysyl isopeptide bond. These results illustrate that specificity for acyl-donor substrates with aromatic ester functions is broader than had previously been supposed and demonstrate that the enzyme can generate products with novel scaffolds.

In this study, we explored the substrate specificity of guinea pig liver TGase by assessing its reactivity toward a variety of potential acyl-donor substrates having an aromatic ester function on the α-carboxyl group of various amino acids (Figure 1D). As well, we assessed its reactivity toward a variety of amino acid derivatives as acyl-acceptor substrates. We observed that the wild-type enzyme can react directly on the α-carboxyl group of *N*-carbobenzyloxyglycine coumarin-7-yl ester (Cbz-Gly-7HC, Figure 1D), demonstrating for the first time that guinea pig liver TGase has an intrinsic peptide synthase activity as it can catalyze the formation of peptide bonds between Gly and a series of small polar and hydrophobic amino acid derivatives.

Furthermore, we performed semi-random mutagenesis of guinea pig liver TGase to further broaden its specificity. Five libraries of mutants were thus created by performing saturation and often combinatorial mutagenesis at conserved active-site residues (Trp241; His305 and Asp306; Trp332; Thr360, Gln362 and Glu363; Tyr519 and Asn520). To determine the activity of >3000 mutants from the five libraries, we developed a rapid

microtiter plate-based screening assay. Using this assay, we identified a mutant with altered specificity for the acyl-acceptor substrate. Single mutant Trp332Phe recognized L-Phe and L-Tyr derivatives as acyl-acceptor substrates while the wild-type enzyme did not, demonstrating that semi-random mutagenesis of the active-site rapidly yielded changes in acyl-acceptor substrate specificity.

6.3 Materials and Methods

6.3.1 Materials

All reagents used were of the highest available purity. Restriction and DNA modifying enzymes were purchased from New England Biolabs (Mississauga, ON) and MBI Fermentas (Burlington, ON). Cellytic B buffer, lysozyme, *N*-α-acetyl-L-lysine methyl ester hydrochloride (*N*-AcLysOMe), glycinamide ($GlyNH_2$) and L-amino acid methyl ester hydrochlorides (ValOMe, SerOMe, PheOMe, TyrOMe, TrpOMe and ProOMe) were purchased from Sigma-Aldrich (Oakville, ON). L-phenylalaninamide ($PheNH_2$) and L-alaninamide hydrochlorides ($AlaNH_2$) were purchased from Novabiochem (Mississauga, ON). QIAPREP plasmid purification kit, QIAEXII agarose gel extraction kit and Ni-NTA agarose resin were purchased from Qiagen (Mississauga, ON). Synthetic oligonucleotides were obtained from Alpha DNA (Montreal, QC) and Integrated DNA Technologies (Coralville, IA). Infra-red dye-labelled sequencing primers were from Li-Cor Biotechnology Division (Lincoln, NB). All aqueous solutions were prepared using water purified with a Millipore BioCell system. DNA sequencing was performed using a kit for Sequenase version 2.0 DNA polymerase (Amersham–Pharmacia Biotech) with a Li-Cor IR2 automated system.

6.3.2 Synthesis of Cbz-Gly-7HC, Cbz-Ala-7HC and Cbz-Phe-7HC

Cbz-Gly-7HC was synthesized based on a previously reported protocol [13]. Namely, 0.2 g (1 mmol) of CbzGly and 0.4 g (2.5 mmol) of 7-hydroxycoumarin (7HC) were dissolved in 10 mL of ethyl acetate. Then 0.22 mL (0.2 g, 2 mmol) of *N*-methylmorpholine and 0.8 mL (0.63 g, 5 mmol) of *N*,*N*-diisopropylcarbodiimide were added with stirring at room temperature. Stirring was continued until the complete disappearance of CbzGly, as followed by thin layer chromatography (ethyl acetate). The reaction mixture was then washed once with 1 M NaOH, three times with 0.1 M NaOH, 3 times with 0.1 M HCl, once with saturated $NaHCO_3$ and once with brine. The organic phase was then dried over $MgSO_4$, filtered and evaporated under reduced pressure. The resulting residue was purified by silica gel chromatography (ethyl acetate) to remove traces of diisopropylurea, giving the desired ester in 70 % yield (0.25 g). Cbz-Ala-7HC and Cbz-Phe-7HC were synthesized according to a similar protocol.

^{1}H NMR (300 MHz, $CDCl_3$) δ4.30 (2H, d), 5.18 (2H, s), 5.33 (1H, s), 6.44 (1H, d), 7.08 (1H, d), 7.10 (1H, s), 7.37 (5H, m), 7.51 (1H, d), 7.70 (1H, d). ^{13}C NMR (75 MHz, $CDCl_3$) δ 168.5, 160.6, 156.7, 154.9, 152.9, 143.1, 136.3, 129.0, 128.9, 128.6, 128.5, 118.4, 117.2, 116.6, 110.5, 67.7, 43.2. HRMS (FAB) calculated for $C_{19}H_{16}NO_6$ ($[M+H]^+$): 354.0972, found 354.0968. Cbz-Ala-7HC: ^{1}H NMR (300 MHz, $CDCl_3$) δ1.63 (3H, d), 4.62 (1H, m), 5.18 (2H, s), 5.33 (1H, d), 6.42 (1H, d), 7.07 (1H, d), 7.09 (1H, s), 7.39 (5H, m), 7.51 (1H, d), 7.70 (1H, d). ^{13}C NMR (75 MHz, $CDCl_3$) δ 171.5, 160.5, 156.0, 154.9, 153.1, 143.1, 136.4, 129.0, 128.9, 128.6, 128.5, 118.4, 117.2, 116.6, 110.5, 67.5, 50.2, 18.5. HRMS (FAB) calculated for $C_{20}H_{18}NO_6$ ($[M+H]^+$): 368.1129, found 368.1118. Cbz-Phe-7HC: ^{1}H NMR (300 MHz, $CDCl_3$) δ 3.26 (2H, m), 4.91 (1H, m), 5.19 (2H, s), 5.35 (1H, d), 6.93 (1H, m), 7.22 (2H, m), 7.39 (9H, m), 7.68 (1H, d), 7.72 (1H, d). ^{13}C NMR (75 MHz, $CDCl_3$) δ 170.1, 160.5, 156.2, 154.9, 152.9, 143.1, 136.3, 135.4, 129.7,

129.2, 128.9, 128.6, 128.5, 127.9, 118.4, 117.2, 116.6, 110.5, 67.6, 55.4, 38.5. HRMS (FAB) calculated for $C_{26}H_{22}NO_6$ ($[M+H]^+$): 444.1442, found 444.1432.

6.3.3 Overexpression and purification of wild-type and mutant TGases

Wild-type and specific mutated recombinant guinea pig liver TGases of interest were overexpressed and purified from *Escherichia coli* on a 1 L scale according to a protocol developed in our laboratory [16] with the following modifications. After Ni-NTA purification, the eluant was transferred to an Amicon Ultra 15 mL tube (Millipore) with a molecular weight cut-off of 30 kDa and the TGase solution was desalted by centrifugation with 25 mM Tris-acetate buffer (pH 7.0) containing 0.5 mM EDTA. The samples were aliquoted, snap-frozen on dry ice and stored at -80 °C. Typical yields were 1.5-10 mg/L of ~85 % pure protein, as estimated from Coomassie Blue staining following SDS-PAGE.

6.3.4 Kinetic assays

All assays were performed in triplicate. The following solutions were prepared: a standard stock buffer solution (100 mM Mops buffer pH 7.0, 5 mM $CaCl_2$, and 0.05 mM EDTA), a 2 mM solution of acyl-donor substrate Cbz-Gly-7HC in DMF, 200 mM solutions of acyl-acceptor substrates *N*-AcLysOMe, $GlyNH_2$, $AlaNH_2$ and SerOMe in water and 800 mM stock solutions of ValOMe, $PheNH_2$, PheOMe, TyrOMe, ProOMe and TrpOMe in 50 % DMF in water. Prior to performing the assays, a "fluorescence coefficient" was determined daily by measuring the arbitrary fluorescence intensities corresponding to five concentrations of 7-hydroxycoumarin at concentrations ranging from 0.05 to 0.5 μM in 5 % DMF in the stock buffer solution at 25 °C. The value of this "fluorescence coefficient" varied only slightly (<5 %) each day. For the hydrolysis reaction, activity was measured by adding 1-3 μg of the purified enzymes to each well of a TCT Luminescence 96-well microtiter plate (Thermo Electron) containing a 0.5-200 μM solution of the acyl-donor substrate in stock buffer. For the transamidation reaction, the same amount of the purified

enzymes was added to 0.5-80 mM of the acyl-acceptor substrate in stock buffer solution containing 100 µM of acyl-donor substrate Cbz-Gly-7HC. The acyl-acceptor substrate was replaced by water or 50 % DMF in water in the blank. DMF was present at 5-7.5 % in the final reaction mixtures. The increase in fluorescence due to the release of 7-hydroxycoumarin was followed against a blank at λ_{ex} 340 nm and λ_{em} 465 nm in a FluoStar Optima microtiter plate reader (BMG Labtech). Linear slopes of fluorescence versus time were measured over the first <10 % conversion of substrate to product and were converted into initial rates using the fluorescence coefficient.

6.3.5 Construction of the libraries of mutants

Five distinct guinea pig liver TGase mutant libraries were created by site-overlap extension PCR [17] by mutating the following conserved active-site residues: Trp241 (library 1); His305 and Asp306 (library 2); Trp332 (library 3); Thr360, Gln362 and Glu363 (library 4) as well as Tyr519 and Asn520 (library 5). The template used was plasmid pQE32-GTG [18] that contains the wild-type guinea pig liver TGase gene with a *Sal*I restriction site at 799 bp and an *Age*I restriction site at position 1728 bp (numbering according to the origin of replication of plasmid pQE32-GTG). All mutations were introduced using primers containing the degenerate "NNS" codon (where N = A, C, G or T and S = C and G) that encodes all twenty amino acids in addition to the amber stop codon. The amplified fragments spanned the portion of the TGase sequence included between the unique *Sal*I and *Age*I restriction sites of the template. They were generated with the external primers GTG-*Sal*I-799-F (5′-TGCTCGCGCCGCAGTCGACCTGTCT-ACGTG-3′) and GTG-*Age*I-1728-R (5′-GTTGGTGCTGCACACCGGTCCCAGGACT-CCATT-3′) containing the *Sal*I and *Age*I restriction sites (underlined), respectively. The external primers were used in combination with sets of complementary pairs of degenerate oligonucleotides that introduced mutations for libraries 1 to 4 (only the coding strands are shown): Library 1 (GTG-W241-NNS: 5′-GCTTCAGGGACGC**NNS**GACAACAAC-

TACAGT-3′), Library 2 (GTG-H305-D306-NNS: 5′-AACTTTAACTCA-GCC**NNSNNS**CAGAACAGCAACCT-3′), Library 3 (GTG-W332-NNS: 5′-AGAGCGAGATGATC**NNS**AACTTCCACTGCTGG-3′) and Library 4 (GTG-LoopIn: 5′-AGGCCCTGGACCCC**NNS**CCC**NNSNNS**AAGAGTGAAGGGACA-TA-3′). For library 4, a previous step of excision of four codons at positions 360-363 (Thr360, Pro361, Gln362 and Glu363) of the guinea pig liver TGase gene was performed using a complementary pair of oligonucleotides (only the coding strand is shown): GTG-LoopOut (5′-CTGGACCCCAAGAGTGAAGGGACATACT-3′). For library 5, the mutations were introduced into the TGase gene using a direct PCR amplification with primer GTG-*Sal*I-799-F and the degenerate oligonucleotide GTG-Y519-N520-NNS-R (5′-CACACA<u>ACCGGT</u>CCCAGGACTCC**SNNSNN**GCTGACGATGCGT-3′, *Age*I restriction site underlined).

The full-length *Sal*I/*Age*I amplicons were digested with *Sal*I/*Age*I, isolated by agarose gel electrophoresis and extracted. The fragment libraries were ligated into pQE32-GTG expression vector *via Sal*I/*Age*I yielding libraries of pQE32-GTG plasmids expressing the mutant TGases. The libraries were transformed by electroporation into *E. coli* SS320. An average of 5×10^4 colonies per library was obtained. Library quality was evaluated by DNA sequencing of 16 to 32 colonies per library following propagation on LB medium containing 100 μg/mL of ampicillin.

Following library validation, all colonies from a given library were pooled in 1 mL fresh LB medium supplemented with ampicillin and were snap-frozen at -80 °C following addition of 500 μL of a sterile 50 % glycerol 1 % NaCl solution. The frozen stocks were used to inoculate cultures for the extraction and purification of the pQE32-GTG plasmids harbouring the mutant TGase libraries. The purified plasmids were transformed into *E. coli* XL-1 Blue cells containing the plasmid pDnaKJ [19] that encodes the molecular chaperones DnaK and DnaJ. An average of ~10^4 colonies per library was obtained.

Colonies were picked into individual wells of Costar 3357 polypropylene 96-well microtiter plates (Corning) containing 100 µL of medium (LB with 100 µg/mL ampicillin and 30 µg/mL chloramphenicol; no ampicillin for the negative control). In each plate, well A1 contained a positive control (*E. coli* XL-1 Blue/pDnaKJ/pQE32-GTG expressing the wild-type guinea pig liver TGase) while well A2 contained a negative control (*E. coli* XL-1 Blue/pDnaKJ). Two plates each for libraries 1 and 3 (188 clones per library) as well as ten plates each for libraries 2, 4 and 5 (940 clones per library) were thus prepared. The plates were covered with a sterile Breathe-Easy gas permeable sealing membrane (Sigma) and incubated overnight at 37 °C with shaking at 250 rpm. After incubation, 50 µL of a sterile 50 % glycerol 1 % NaCl solution was added to all wells and each plate was covered with a Storage Mat III (Corning) previously sterilized in ethanol. The resulting "mother" plates were snap-frozen and stored at -80 °C.

6.3.6 Library representation calculations

Library representation was calculated based on a Poisson distribution, as follows. The libraries were designed to encode n possible different theoretical variants at the DNA level (Table 3, number of encoded genes; some redundancy exists at the protein level). Library screening entailed sampling m colonies, randomly, from this theoretical pool; we thus formed experimental samples of m variants per library. We calculated λ, the expected number of missing theoretical variants within the m variants: $\lambda = n\,(1-{}^{1}/_{n})^{m}$ [20]. Library representation is given as $1 - ({}^{m}/_{n})$ and is expressed as a percentage.

6.3.7 Preparation of the crude bacterial lysate for screening

The mother plates containing the libraries were used to inoculate sterile, round bottom Costar 3799 (Corning) 96-well microtiter plates ("daughter" plates) containing 100 µL of LB per well, supplemented with the appropriate antibiotics. The daughter plates were incubated overnight (37 °C, 250 rpm). The bacterial cultures (5 µL) were used to

inoculate 100 µL of fresh medium in new 96-well microtiter plates. The plates were incubated (37 °C, 250 rpm) for 3 h, at which time protein expression was induced by addition of 1 mM isopropyl-β-D-thiogalactoside with overnight incubation (30 °C, 250 rpm).

For cell lysis, the plates were snap-frozen on crushed dry ice and then thawed at 37 °C. Lysis buffer (20 µL of a 6 mg/mL lysozyme in Cellytic B buffer solution) was added to each well. The plates were incubated at 30 °C for 30 min. The crude bacterial lysate was used directly in the screening assay.

6.3.8 Screening assay

A screening procedure was developed to measure the activity of the five mutant libraries. This procedure utilizes Cbz-Xaa-7HC esters (Figure 1D) as acyl-donor substrates. A standard stock buffer solution (100 mM Mops buffer pH 7.0, 5 mM $CaCl_2$, and 0.05 mM EDTA), 1.5 mM stock solutions of acyl-donor substrates Cbz-Gly-7HC and Cbz-Ala-7HC in DMF and 200 mM stock solutions of acyl-acceptor substrates *N*-AcLysOMe in water and $PheNH_2$ in 50 % DMF were prepared. Transamidation activity was measured by adding 20 µL of the crude bacterial lysates to 180 µL of reaction buffer containing 75 µM of the acyl-donor substrate and 10 mM of the acyl-acceptor substrate in the stock buffer solution in each well of a TCT Luminescence 96-well microtiter plate (Thermo Electron). Water replaced acyl-acceptor substrate for the hydrolysis reaction. DMF was present at 5-7.5 % in the final reaction mixtures. The increase in fluorescence due to the release of 7-hydroxycoumarin was followed at λ_{ex} 340 nm and λ_{em} 465 nm in a HTS 7000 plate reader (Perkin Elmer). Linear slopes of fluorescence versus time were measured over 15 min.

6.4 Results

6.4.1 Cbz-Gly-7HC is hydrolyzed by wild-type guinea pig liver TGase

We previously demonstrated that guinea pig liver TGase does not require the L-Gln or L-Glu scaffold in its acyl-donor substrates when the leaving group is an aromatic alcohol: the enzyme reacted with Cbz-Gly-GABA-7HC as well as Cbz-Phe-GABA-7HC (Figure 1C) with K_M values of 7-9 μM and k_{cat} values of 45-75 min^{-1} [13]. To verify if guinea pig liver TGase could react with analogs of these compounds having the 7-hydroxycoumarin leaving group directly bound to the α-carboxyl group of an amino acid, *N*-carbobenzyloxyamino acid coumarin-7-yl esters with substituents of varying bulk (Cbz-Gly-7HC, Cbz-Ala-7HC and Cbz-Phe-7HC, Figure 1D) were synthesized and their reactivity was evaluated in the hydrolysis reaction. Michaelis-Menten kinetics demonstrated that only compound Cbz-Gly-7HC is an acyl-donor substrate of wild-type guinea pig liver TGase, having a K_M of 14 ± 3 μM and a k_{cat} of 6.8 ± 0.5 min^{-1} (Table 1). This K_M is similar to that of Cbz-Gly-GABA-7HC (9 ± 2 μM) while the k_{cat} is ~10-fold lower (75 ± 5 min^{-1}) [13]. The lower k_{cat} appears to be due to the absence of the γ-aminobutyric acid (GABA) linker in Cbz-Gly-7HC relative to Cbz-Gly-GABA-7HC, which may result in a decreased accessibility of the substrate reactive carbonyl group for nucleophilic attack by the catalytic thiol. An alternative explanation would be that the absence of the GABA linker hinders hydrolysis of the acyl-enzyme intermediate formed during the reaction by modifying accessibility or nucleophilicity of the acceptor water molecule. Cbz-Ala-7HC did not react with guinea pig liver TGase and Cbz-Phe-7HC was not soluble enough to test against guinea pig liver TGase. The lack of reactivity of Cbz-Ala-7HC appears to result from the methyl side chain of the Ala residue that may hinder productive binding at the active-site.

Table 1. Kinetic constants of active guinea pig liver TGase mutants for hydrolysis of Cbz-Gly-7HC

TGase	K_M[a] (μM)	k_{cat}[a] (min^{-1})	k_{cat}/K_M ($min^{-1}\mu M^{-1}$)	k_{cat}/K_M WT / k_{cat}/K_M mutant
Wild-type	14 ± 3	6.8 ± 0.5	0.49	1.0
Library 3				
Trp332Phe	24 ± 3	7.3 ± 0.4	0.30	0.6
Trp332Tyr	11 ± 1	7.6 ± 0.2	0.69	1.4
Library 5				
Tyr519Val	23 ± 4	4.3 ± 0.2	0.19	0.4
Asn520Asp	18 ± 2	4.5 ± 0.1	0.25	0.5
Tyr519Phe + Asn520Leu	16 ± 3	1.26 ± 0.05	0.08	0.2
Tyr519Cys + Asn520Gln	27 ± 3	5.6 ± 0.2	0.21	0.4
Tyr519His + Asn520Arg	25 ± 4	8.9 ± 0.7	0.36	0.7
Tyr519Gly + Asn520Ser	17 ± 2	4.8 ± 0.1	0.28	0.6
Tyr519Ser + Asn520Ser	16 ± 2	0.90 ± 0.04	0.06	0.1

[a] Kinetic parameters of acyl-donor substrate Cbz-Gly-7HC.

6.4.2 Wild-type guinea pig liver TGase can catalyse the formation of peptide bonds

The kinetic parameters for the guinea pig liver TGase-catalyzed transamidation reaction of acyl-donor substrate Cbz-Gly-7HC and acyl-acceptor substrate *N*-AcLysOMe were calculated as the best-fit values using nonlinear regression of the Michaelis-Menten equation $v_0 = v_{max}$ [S] / (K_M + [S]). It should be noted that the k_{cat} and K_M values reported for all transamidation reactions in this study are apparent k_{cat} and K_M values determined at a high concentration of donor substrate Cbz-Gly-7HC equal to ~7 times the K_M value. An

increase in the rate of liberation of 7-hydroxycoumarin from Cbz-Gly-7HC in the presence of the widely used acyl-acceptor substrate *N*-AcLysOMe was observed. Wild-type guinea pig liver TGase catalyzes the formation of a covalent bond between the α-carboxyl group of glycine in Cbz-Gly-7HC and the ε-amino group of lysine in *N*-AcLysOMe with K_M and k_{cat} values of 4 ± 1 mM and 12 ± 1 min^{-1}, respectively (Table 2). As tissue TGase can use amino acid derivatives as acyl-acceptor substrates [9], we tested a variety of amino acid derivatives for their potential as acyl-acceptor substrates. The bulky aromatic amino acid derivatives $PheNH_2$, PheOMe, TyrOMe and TrpOMe were not substrates of wild-type guinea pig liver TGase; neither was ProOMe (data not shown). This correlates with previous observations that amines with adjacent bulky substituents are not acyl-acceptor substrates [9]. However, the small hydrophobic amino acid derivatives $GlyNH_2$ and $AlaNH_2$ acted as substrates. $GlyNH_2$ had a K_M of 1.9 ± 0.3 mM and a k_{cat} of 11.0 ± 0.4 min^{-1} while $AlaNH_2$ had a K_M of 4 ± 1 mM and a k_{cat} of 4.2 ± 0.2 min^{-1}. $GlyNH_2$ and $AlaNH_2$ have been identified previously as acyl-acceptor substrates of guinea pig liver TGase [9]. ValOMe had an inhibitory effect at concentrations higher than 10 mM. SerOMe, a small polar amino acid derivative, was also found to serve as a substrate of the enzyme with a K_M of 5 ± 2 mM and a k_{cat} of 11 ± 1 min^{-1}. These results demonstrate that wild-type guinea pig liver TGase can catalyze the formation of Cbz-Gly-Xaa dipeptides (where Xaa is Gly, Ala or Ser) through a peptide synthase activity.

Table 2. Kinetic constants of active guinea pig liver TGase mutants for transamidation reaction of Cbz-Gly-7HC and *N*-AcLysOMe

TGase	K_M[a] (mM)	k_{cat}[a] (min^{-1})	k_{cat}/K_M ($min^{-1}mM^{-1}$)	k_{cat}/K_M WT / k_{cat}/K_M mutant
Wild-type	4 ± 1	12 ± 1	3.0	1.0
Library 3				
Trp332Phe	0.8 ± 0.2	9.2 ± 0.5	12	4.0
Trp332Tyr	14 ± 5	24 ± 4	1.7	0.6
Library 5				
Tyr519Val	2.7 ± 0.7	4.9 ± 0.3	1.8	0.6
Asn520Asp	3 ± 1	6.3 ± 0.5	2.1	0.7
Tyr519Phe + Asn520Leu	5 ± 1	4.1 ± 0.3	0.8	0.3
Tyr519Cys + Asn520Gln	6 ± 1	9.5 ± 0.7	1.6	0.5
Tyr519His + Asn520Arg	4 ± 2	13 ± 1	3.3	1.1
Tyr519Gly + Asn520Ser	4 ± 1	6.6 ± 0.5	1.7	0.6
Tyr519Ser + Asn520Ser	12 ± 3	3.4 ± 0.2	0.3	0.1

[a] Kinetic parameters of acyl-acceptor substrate *N*-AcLysOMe. Note that for values reported here, $k_{cat} = k_{cat}^{app}$ and $K_M = K_M^{app}$.

6.4.3 Semi-random mutagenesis of guinea pig liver TGase

Semi-random mutagenesis of the active-site of guinea pig liver TGase was undertaken to broaden its specificity for the acyl-donor and acyl-acceptor substrates in order to generate mutants capable of catalyzing the formation of a wider variety of dipeptides. Five mutant libraries were created by mutating the following conserved active-site residues (Table 3, Figure 2): Trp241; His305 and Asp306; Trp332; Thr360, Gln362 and Glu363 as well as Tyr519 and Asn520. These residues were selected because they are

within 5 Å of the L-Gln side chain of the acyl-donor substrate Cbz-Gln-Gly in the model of tissue TGase with substrate bound at the active-site [21], and thus may be implicated in ligand binding. These residues are not considered to be directly implicated in catalysis. All 20 amino acids were encoded at each position using the degenerate "NNS" codon. Thus, for saturation libraries 1 (Trp241) and 3 (Trp332), 32 codons were introduced that encode 20 amino acids. Libraries 2 (His305 and Asp306), 4 (Thr360, Gln362 and Glu363) and 5 (Tyr519 and Asn520) are combinatorial: libraries 2 and 5 theoretically contain 1024 combinations of codons encoding 400 different proteins while library 4 contains 32768 combinations of codons encoding 8000 different proteins.

Table 3. Libraries of mutant guinea pig liver TGases generated in this study

Library	Mutated residue(s)	Encoded genes	Encoded proteins	Clones screened
1	Trp241	32	20	188
2	His305 and Asp306	1024	400	940
3	Trp332	32	20	188
4	Thr360, Gln362 and Glu363	32768	8000	940
5	Tyr519 and Asn520	1024	400	940

Library quality was assessed by DNA sequencing prior to screening. Nucleotide representation for each degenerate codon generally followed the expected statistical distribution, and no non-specific mutations were observed. Transformation into bacterial cells for expression yielded ~10^4 colonies for each library, amply satisfying our relatively limited throughput: a maximum of 940 clones per library were screened using the microtiter plate assay (Table 3).

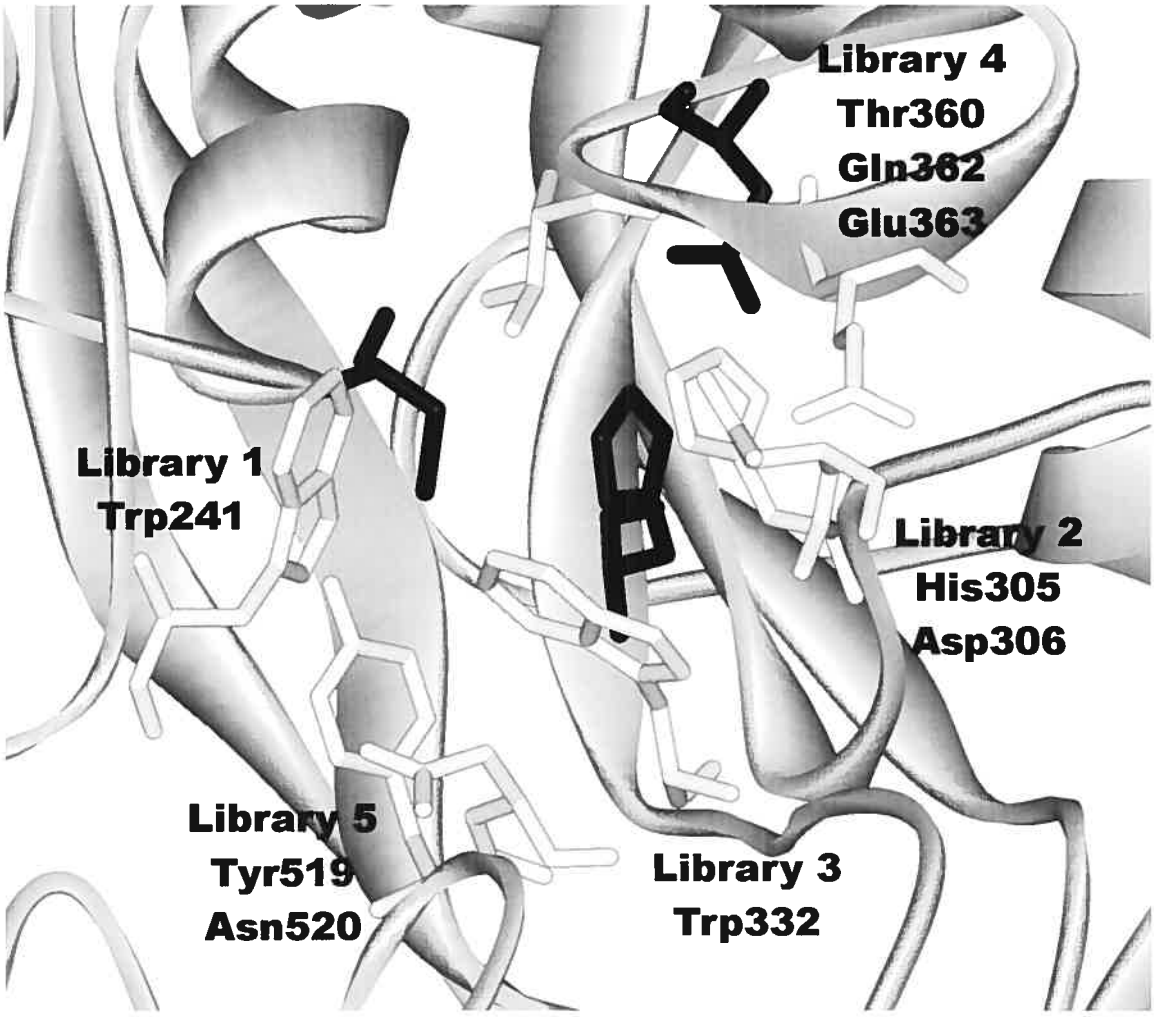


Figure 2. Active-site residues of guinea pig liver TGase [18]. Catalytic residues Cys277, His335 and Asp358 are in black. Residues from the five libraries are in white.

6.4.4 Library screening

To detect active mutants towards the formation of peptide bonds, a fluorescence-based microtiter plate screening assay was developed. In this assay, the release of 7-hydroxycoumarin that occurs in the first step of the TGase mechanism, the acylation step, is measured. Thus, this assay measures the steady-state rate of hydrolysis and/or

transamidation. In each case, TGase deacylation is the rate-limiting step [11, 22]. 188 clones of libraries 1 and 3 were screened, approximately 6 times the library size (32 DNA variants) giving a representation of >99 % of the whole library. For libraries 2, 4 and 5, 940 clones were screened, resulting in approximately 60 % library coverage for libraries 2 and 5 (1024 DNA variants) and <3 % for library 4 (32768 DNA variants). Library representation was calculated as described under Methods. It should be noted that this screening assay is also easily amenable to automation for screening larger numbers of mutants in a library versus library approach (mutants versus commercially-available acyl-acceptor substrates).

For each library, we first measured the rate of hydrolysis using Cbz-Gly-7HC or Cbz-Ala-7HC as acyl-donor substrates. As noted above, Cbz-Gly-7HC is a substrate of wild-type guinea pig liver TGase. It was used to detect active mutants among the libraries. Cbz-Ala-7HC is not a substrate of wild-type guinea pig liver TGase and was used to screen for mutants with a different specificity for the acyl-donor substrate. Cbz-Phe-7HC was not used for screening due to its low aqueous solubility (<50 μM).

None of the >3000 clones screened were capable of hydrolyzing Cbz-Ala-7HC at a detectable rate. Thus, the semi-random mutagenesis performed was not consistent with this particular modification of enzyme specificity for the acyl-donor substrate. However, library screening with Cbz-Gly-7HC in the presence or absence of acyl-acceptor *N*-AcLysOMe yielded active mutants from libraries 3 and 5 for the hydrolysis and transamidation reactions. DNA sequencing of the active mutants (Table 4) revealed that library 3 (single mutation) yielded only two active mutants (Trp332Phe and Trp332Tyr), suggesting that neutral aromatic amino acids are the only residues that can be introduced at this position to give a detectable activity. Library 5 (double mutations) yielded 20 active mutants, three of which were single mutants. Although residues Tyr519 and Asn520 are conserved in nature, they are more permissive to mutations than the other positions screened as a variety of hydrophobic, uncharged polar and charged amino acids were identified at each position.

Table 4. Active mutants identified by screening for transamidation of Cbz-Gly-7HC and *N*-AcLysOMe

	Residue 332	Residue 519	Residue 520
Wild-type	Trp	Tyr	Asn
Library 3	*Phe*[a]		
	Tyr		
Library 5		*Ser*	*Thr*
		Ser	*Ser*
		Ala	*Gln*
		Thr	*Lys*
		Gln	*Val*
		Tyr	*Asp*
		Cys	*Gln*
		His	*Arg*
		Val	Asn
		Asn	*Arg*
		Phe	*Leu*
		Asn	*Arg*
		Gly	*Glu*
		Cys	Asn
		Cys	*Arg*
		Ser	*Lys*
		Asn	*Gly*
		Thr	*Val*
		Gly	*Ser*
		Pro	*Ser*

[a] Mutations are indicated in italics.

Indeed, 11 out of 20 encoded amino acids were identified at each position (Gly, Ala, Val, Pro, Ser, Thr, Cys, Asn, Gln, Tyr and His for position 519 and Gly, Val, Leu, Ser, Thr, Asn, Gln, Asp, Glu, Lys and Arg for position 520) demonstrating the high tolerance to sequence variation at these positions. An even higher variability of amino acids would have been expected at these positions if our library screening had been exhaustive. Throughout screening, the wild-type enzyme was also identified as active. No active mutants from libraries 1 (Trp241), 2 (His305 and Asp306) and 4 (Thr360, Gln362 and Glu363) were identified.

Screening of the five mutant libraries using acyl-donor Cbz-Gly-7HC with the potential acyl-acceptor $PheNH_2$ was also performed. As $PheNH_2$ is not an acyl-acceptor substrate of wild-type guinea pig liver TGase, this screening was performed to identify mutants with altered specificity. However, this screening did not allow us to detect mutants that had an increased rate of transamidation relative to hydrolysis in the presence of $PheNH_2$, potentially due to our non-exhaustive screening and to the low sensitivity of our assay that allows only mutants with near-native activity to be identified.

6.4.5 Enzyme kinetics of identified active mutants

Kinetic constants for acyl-donor substrate Cbz-Gly-7HC in the hydrolysis reaction and acyl-acceptor substrate *N*-AcLysOMe in the transamidation reaction were determined for all the active mutants from library 3 and for the seven most active mutants from library 5 identified by screening (Tables 1 and 2). For the hydrolysis reaction, all mutants except mutant Trp332Tyr exhibited a ~1.5- to 10-fold decrease in catalytic efficiency. Mutant Trp332Tyr was slightly more catalytically efficient than the wild-type enzyme. For the transamidation reaction with *N*-AcLysOMe, all mutants except single mutant Trp332Phe and double mutant Tyr519His+Asn520Arg, had a ~1.5- to 10-fold decrease in catalytic efficiency, demonstrating a good correlation between the relative catalytic efficiency of hydrolysis and transamidation for most mutants. Double mutant

Tyr519His + Asn520Arg was found to be as catalytically efficient as the wild-type enzyme with respect to transamidation.

The mutations at position 332 from the two identified active mutants resulted in different effects. Whereas mutant Trp332Tyr was more efficient at catalyzing the hydrolysis of Cbz-Gly-7HC than the wild-type, it was less efficient than the wild-type enzyme for catalysis of the transamidation reaction with *N*-AcLysOMe. However, mutant Trp332Phe showed an impressive 4-fold increase in catalytic efficiency of transamidation relative to the wild-type while being only slightly less efficient than wild-type for hydrolysis. This suggests that the Trp332Phe mutation improves transamidation efficiency without significantly affecting hydrolytic efficiency. The opposing results for these two mutants are surprising since both display uncharged aromatic amino acids at position 332.

Kinetic parameters for the potential acyl-acceptor substrate $PheNH_2$ were also measured for both active mutants from library 3 and for the seven most active mutants of library 5. Although no mutants showed activity by the plate-based screening method using this compound, mutant Trp332Phe showed significant transamidation activity with $PheNH_2$. The K_M and k_{cat} values determined were 12 ± 5 mM and 1.6 ± 0.2 min^{-1}, respectively (Table 5). The resulting catalytic efficiency value of 0.1 $min^{-1}mM^{-1}$ is 120-fold lower than the corresponding value with acyl-acceptor *N*-AcLysOMe (12 $min^{-1}mM^{-1}$; Table 2) and is apparently below the limit of detection of the plate-based screening assay. The absolute transamidation efficiency of mutant Trp332Phe relative to that of the wild-type likely accounts for it being the only mutant with detectable $PheNH_2$ transamidation. Mutant Trp332Phe also recognized PheOMe with similar K_M and k_{cat}, demonstrating that the C-terminal amide or methyl ester functional groups of the L-Phe derivatives have no significant effect on the kinetic parameters. This is in sharp contrast with the charged carboxylate group that renders amino acids unreactive as acyl-acceptor substrates [9].

Table 5. Acyl-acceptor substrate specificity of wild-type guinea pig liver TGase and mutant Trp332Phe

	Wild-type			Trp332Phe		
Acyl-acceptor substrate	K_M[a] (mM)	k_{cat} (min^{-1})	k_{cat}/K_M (min^{-1} mM^{-1})	K_M (mM)	k_{cat} (min^{-1})	k_{cat}/K_M (min^{-1} mM^{-1})
$GlyNH_2$	1.9 ± 0.3	11.0 ± 0.4	5.8	1.7 ± 0.3	9.7 ± 0.5	5.8
$PheNH_2$	Not recognized			12 ± 5	1.6 ± 0.2	0.1
PheOMe	Not recognized			15 ± 6	2.1 ± 0.2	0.1
TyrOMe	Not recognized			3 ± 1	2.3 ± 0.2	0.8

[a] Note that for values reported here, $k_{cat} = k_{cat}^{app}$ and $K_M = K_M^{app}$.

TyrOMe was also found to be an acyl-acceptor substrate of mutant Trp332Phe. While its k_{cat} was again similar to that of $PheNH_2$ or PheOMe, its K_M was 4-fold lower (Table 5). This suggests that the hydroxyl group of Tyr provides an additional beneficial interaction with the acyl-enzyme intermediate, improving productive binding. The catalytic efficiency for the acyl-acceptor substrate glycinamide ($GlyNH_2$) is the same for mutant Trp332Phe and the wild-type (Table 5), suggesting that the mutation does not interfere with catalysis of transamidation with small primary amines as acyl-acceptor substrates. Thus, the mutant does not lose reactivity toward substrates recognized by the wild-type enzyme while exhibiting a broadened specificity.

6.5 Discussion

6.5.1 Wild-type guinea pig liver TGase has a peptide synthase activity

Herein, we report for the first time the intrinsic peptide synthase activity of wild-type guinea pig liver TGase. Our results demonstrate that Cbz-Gly-7HC is an acyl-donor

substrate of wild-type guinea pig liver TGase and that various small amino acid derivatives are acyl-acceptor substrates. Aromatic esters acting as acyl-donor substrates of guinea pig liver TGase has been previously demonstrated [11, 13]. Furthermore, Clarke *et al.* [9] reported that glycinamide-N^{15} can serve as an acyl-acceptor substrate for guinea pig liver TGase, proceeding with retention of the amide N^{15} atom. This indicates that formation of the covalent bond between glycinamide-N^{15} and the L-Gln residue of the acyl-donor substrate occurred through the α-amino group of glycinamide rather than through the amide function. These previously reported observations provide insight for interpreting the peptide synthase activity of wild-type guinea pig liver TGase reported herein.

The peptide synthase activity of guinea pig liver TGase we observed results from its transferase activity: the enzyme can transfer an α-Gly moiety onto the α-amino group of various small amino acid derivatives, thus forming peptide bonds. This transferase activity of TGases relies on their capacity to exclude water from the active-site [23]. If water has free access to the thioester bond of the covalent acyl-enzyme intermediate, amine acyl-acceptor substrates would not be able to compete with it for the acyl transfer reaction, as water is much more abundant. Thus, the intermediate thioester must be sequestered in the active-site long enough for amines to enter and act as acyl-acceptor substrates. It is believed that the exclusion of water results from hydrophobic residues forming a tunnel leading to the catalytic residues. These hydrophobic residues have been identified as residues Trp241, His305 and Trp332 which form the walls of the active-site tunnel [21]. This ability to exclude water from the active-site differentiates TGases from the cysteine proteases, such as papain, that share a similar segment of α-helix and β-sheet containing an identical catalytic triad [24].

Proteases are known to catalyze the synthesis of peptides under certain conditions. In a thermodynamically-controlled process, amino acid condensation is favoured over peptide hydrolysis by shifting equilibria through product precipitation or decreasing water activity. In a kinetically-controlled process, the aminolysis of an acyl-enzyme formed from

an activated acyl-donor substrate is favoured over its hydrolysis by using high concentrations of acyl-acceptor substrate. The kinetic process requires a protease such as serine or cysteine proteases that can form a covalent acyl-enzyme intermediate [25], as is the case with TGases. Competition between hydrolysis and aminolysis is always present during the degradation of this acyl-enzyme intermediate, resulting in lower yields for the synthesis of peptides, since proteases are not efficient at excluding water from the active-site – rather, they are designed to bind and activate water as part of their native catalytic cycle. Furthermore, proteases can hydrolyse their peptide products, further lowering the overall yield of synthetic peptide.

Proteases used in kinetically-controlled peptide synthesis have a transamidation/hydrolysis ratio in the range of 10^2-10^4 [26, 27] whereas guinea pig liver TGase has a similar 10^2 increase in rate of transamidation relative to hydrolysis when the acceptor substrate is hydroxylamine [11]. The catalytic efficiency of wild-type guinea pig liver TGase for the synthesis of the Cbz-Gly-GlyNH_2 and Cbz-Gly-AlaNH_2 dipeptides is 5.8 and 1.1 $min^{-1}mM^{-1}$, respectively. Papain, a cysteine protease that has been used in peptide synthesis, has catalytic efficiencies of 0.3 and 2.9 $min^{-1}mM^{-1}$ for the synthesis of the Boc-Gly-Phe-N_2H_2Ph dipeptide and the Boc-Tyr(Bzl)-Gly-Gly-Phe-Leu-N_2H_2Ph pentapeptide, respectively [28]. These catalytic efficiencies are within the same order of magnitude than those of wild-type guinea pig liver TGase, suggesting that TGase could be used as a catalyst for the synthesis of peptide bonds. An advantage of guinea pig liver TGase-catalyzed peptide synthesis is that it requires no organic co-solvent. Indeed, the 5 % DMF used in the transamidation assay of Cbz-Gly-7HC and GlyNH_2 by TGase is required only to help solubilize the acyl-donor substrate. This is not the case with papain, which catalyzed the synthesis of the Boc-Tyr(Bzl)-Gly-Gly-Phe-Leu-N_2H_2Ph peptide in a mixture containing 40 % ethanol [29] to decrease the activity of water, since hydrolysis is the preferred, native reaction of the protease. A further advantage of TGase for the synthesis of peptide bonds is that TGase cannot recognize secondary amides as acyl-donor substrates, thus limiting hydrolysis of the dipeptide product and potentially increasing yields.

We tested wild-type guinea pig liver TGase with the amide analog of Cbz-Gly-7HC, Cbz-glycinamide (Cbz-GlyNH_2), to verify if it could act as an acyl-donor substrate. This compound does not react with TGase at concentrations up to 50 mM. The K_M value for this compound is apparently greater than its solubility limit. It has previously been observed that aromatic ester acyl-donor substrates of TGase have a lower K_M value than the corresponding amide: Cbz-Gln-Gly (Figure 1A) has a K_M of 3.2 mM [21] whereas its aromatic ester analog Cbz-Glu(γPNP)-Gly (Figure 1B) has a K_M of 0.02 mM [11]. The 160-fold lower K_M of the aromatic ester could be due to improved binding conferred by the *p*-nitrophenol aromatic leaving group. This improved binding could also occur in Cbz-Gly-7HC relative to Cbz-GlyNH_2 through beneficial π-stacking interactions between the aromatic leaving group and the aromatic side chains of the three tunnel-wall residues of guinea pig liver TGase (Trp241, His305 and Trp332). This apparent need of an activated leaving group in acyl-donor substrates for peptide synthesis by native guinea pig liver TGase represents a serious limitation for their application to peptide synthesis. This limitation is not observed for proteases, for example, which can react directly on the carboxylate group of acyl-donors.

6.5.2 Modification of guinea pig liver TGase specificity

As discussed above, guinea pig liver TGase may be useful for peptide synthesis. However, to improve its usefulness, the enzyme specificity should be broadened. Wild-type guinea pig liver TGase cannot recognize Cbz-Ala-7HC as an acyl-donor substrate, therefore preventing it from forming Cbz-Ala-Xaa dipeptides. Furthermore, amino acid derivatives having bulky side-chains are not acyl-acceptor substrates of the enzyme. To further broaden its specificity for acyl-donor and acyl-acceptor substrates, we performed semi-random mutagenesis of conserved active-site residues coupled to screening with desired compounds such as the aforementioned Cbz-Ala-7HC. Our rationale was that modification of the

specificity would be achieved more readily through a semi-random approach by mutating residues that directly interact with the substrate.

Although through this approach we did not identify any mutants with altered acyl-donor specificity of the enzyme, we rapidly identified a point mutant, Trp332Phe, with a different specificity for the acyl-acceptor substrate. This mutant can form peptide bonds between Cbz-Gly and the neutral aromatic amino acids L-Phe and L-Tyr, and represents the first example of a TGase that can utilize an amine with an adjacent bulky substituent as an acyl-acceptor substrate. Indeed, this mutant catalyzed the formation of the Cbz-Gly-PheNH_2, Cbz-Gly-PheOMe and Cbz-Gly-TyrOMe dipeptides. However, the catalytic efficiency of mutant Trp332Phe with these substrates (0.1-0.8 $min^{-1}mM^{-1}$) is one order of magnitude lower than that with acyl-acceptor substrates such as GlyNH_2 and *N*-AcLysOMe (5.8 $min^{-1}mM^{-1}$ and 12.0 $min^{-1}mM^{-1}$, respectively).

Comparison of the catalytic efficiency of guinea pig liver TGase mutant Trp332Phe for the formation of the Gly-Phe peptide bond (0.1 $min^{-1}mM^{-1}$) with the corresponding value for papain (0.3 $min^{-1}mM^{-1}$; Table 6) [28] reveals that they are in the same order of magnitude. Thus, Trp332Phe, while being less efficient at forming Gly-Phe peptide bonds than Gly-Gly peptide bonds (Table 5), is roughly as efficient as papain for the formation of Gly-Phe peptide bonds. This demonstrates that mutant Trp332Phe of guinea pig liver TGase could be useful for peptide synthesis.

Table 6. Comparison of guinea pig liver TGase mutant Trp332Phe and papain for synthesis of Gly-Phe peptide bonds

	Guinea pig liver TGase mutant Trp332Phe[a]	Papain[b]
Acyl-acceptor	$PheNH_2$	$PheN_2H_2Ph$
K_M (mM)	12 ± 5	51.2 ± 4.2[d]
k_{cat}[c] (min^{-1})	1.6 ± 0.2	16.2 ± 1.8[d]
k_{cat}/K_M[c] ($min^{-1} M^{-1}$)	0.1	0.3[d]

[a] Acyl-donor substrate is Cbz-Gly-7HC.

[b] Acyl-donor substrate is Boc-Gly.

[c] Values are for the transamidation reaction where the acyl-donor substrate is at a saturating concentration.

[d] Values are from [28].

Of further import is that the change of specificity for acyl-acceptor substrates observed in mutant Trp332Phe results from a single, conservative mutation. Visualization of the guinea pig liver TGase homology model [18] reveals that residue 332 is located at the entrance of the active-site. The replacement of Trp by Phe should increase the volume of the active-site cavity while retaining the hydrophobic character proposed to be important for activity [21]. The increase in active-site cavity size could allow the entry of the larger acyl-acceptor substrates that appear to be excluded in the wild-type. Further modifications to the active-site cavity could lead to various specificities for the acyl-acceptor substrate.

6.5.3 Structure/function analysis of active-site residues of guinea pig liver TGase

Only libraries 3 and 5 yielded active mutants. Library 3 is a saturation library at position Trp332. Only two active mutants were identified by screening >99 % of all possible mutants. These mutants, Trp332Phe and Trp332Tyr have neutral aromatic acid

substitutions at position 332 indicating that the enzyme requires a bulky aromatic residue at position 332 to maintain its activity, stability or proper folding. Previous experimental evidence supports this. Conserved active-site residue Trp332 of human tissue TGase was previously mutated to Ala or Phe [30]. Mutant Trp332Ala did not allow recovery of the recombinant protein possibly due to improper folding reducing its solubility. Residue Trp332 is located at the entrance of the hydrophobic active-site tunnel and has been proposed to form a "gate" to the active-site [21]. Replacement of this residue with a small residue such as Ala might expose the hydrophobic tunnel and decrease enzyme solubility. Mutant Trp332Phe of human tissue TGase was found to have a transamidation activity similar to wild-type [30]. This observation correlates with our results of the near-native catalytic efficiency of guinea pig liver TGase mutant Trp332Phe. Unfortunately, no kinetic parameters were reported for human tissue TGase mutant Trp332Phe preventing a more thorough comparison with our mutant.

Screening of combinatorial library 5 allowed us to identify 20 mutants with near-native activity. These mutants had 11 out of 20 possible encoded amino acids at either position 519 or 520, with no bias towards polar, hydrophobic, charged or aromatic amino acids, indicating that these positions are very permissive to mutations. Our previous mutagenesis studies of guinea pig liver TGase [18] have demonstrated the permissibility to mutations of conserved active-site residue Tyr519 which was mutated to amino acids with side chains of diverse polarity (Leu, Phe, Gln and His). These mutagenesis experiments demonstrated that Tyr519 is not essential for the hydrolysis or transamidation reactions catalyzed by tissue TGase although it is conserved. Furthermore, mutation of residue Tyr519 in combination with mutation of residue Cys336, which is located next to the catalytic Cys residue, yielded mutants with near-native catalytic efficiency, again implying that the active-site of guinea pig liver TGase is tolerant to mutations at these positions [18]. Thus, it is not surprising that we identified 11 out of 20 possible mutations at position 519 during our screening. Since we only screened 940 mutants, representing approximately

60 % of the library, we could expect to identify more mutants with near-native activity with a more exhaustive screen.

Library 1 yielded no mutants with detectable, near-native hydrolysis or transamidation activities. This library is a saturation library of residue Trp241 and was screened with the same representation of >99 % as library 3. Residue Trp241 of human tissue TGase was previously mutated to various amino acids [31]. It was observed that non-aromatic substitutions at this position (Trp241Gly and Trp241Ala) resulted in mutants with hardly any detectable activity while aromatic substitutions (Trp241Phe, Trp241Tyr and Trp241His) yielded active mutants that nonetheless showed a 25- to 165-fold decrease in catalytic efficiency relative to wild-type. The sensitivity of our screening assay allowed us to detect only mutants displaying near-native activity, therefore increasing its stringency. Mutants having a 25- to 165-fold decrease in catalytic efficiency would not be detected. Thus, although mutants Trp241Phe, Trp241Tyr and Trp241His were present in library 1, we did not identify them as active during our screening.

Library 2 is a combinatorial library at positions His305 and Asp306. It contains the same number of mutants as library 5 and thus our screening should have represented approximately 60 % of the whole library. Unlike library 5, library 2 did not yield mutants with any detectable hydrolysis or transamidation activities, suggesting that conserved residues His305 and Asp306 may play an important role in the stability or activity of the enzyme. These residues have been postulated to be implicated in exclusion of water from the active-site through the formation of a salt bridge between them [23]. Furthermore, these residues have been postulated to play an important role in the positioning and deprotonation of the amine substrates to the active-site [32, 33]. Thus, the lack of active mutants may result from the loss of important interactions between the enzyme and the acyl-acceptor substrates during the deacylation step, which is the step we measure.

Library 4 is a combinatorial library containing 32768 genetic variants encoding for 8000 mutant proteins and was the largest library. It contained mutants at positions Thr360,

Gln362 and Glu363: none of these residues have been mutated previously. Consequently, it is the library with the lowest representation during screening: only 940 clones were screened which represents <3 % of the whole library. No active mutants from this library were identified due to the limited range of our screening. As a result, it is difficult to propose a role for residues Thr360, Gln362 and Glu363. Nonetheless, these residues belong to the same loop as catalytic residue Asp358 (Figure 2). Mutations of this loop might disrupt the proper positioning of this catalytic residue, affecting activity.

6.6 Conclusion

Herein we demonstrate for the first time the peptide synthase activity of wild-type guinea pig liver TGase. This enzyme can form Gly-Xaa peptide bonds between Cbz-protected Gly and a variety of small polar or hydrophobic amino acid derivatives with a similar catalytic efficiency as the cysteine protease papain. Although the specificity of guinea pig liver TGase for peptide bond formation is limited, it could be a new tool for the enzymatic synthesis of peptides and complement the known specificities of other proteases. To expand its specificity, we performed semi-random mutagenesis of active-site residues and rapidly identified mutant Trp332Phe by a fluorimetric microplate-based assay. This mutant has a different specificity for the acyl-acceptor substrate. The semi-random mutagenesis also allowed us to acquire structure/function information on the role of the active-site residues of guinea pig liver tissue TGase.

6.7 References

1 Fesus, L. and Piacentini, M. (2002) Transglutaminase 2: an enigmatic enzyme with diverse functions. Trends in Biochemical Sciences **27**, 534-539

2 Griffin, M., Casadio, R. and Bergamini, C. M. (2002) Transglutaminases: nature's biological glues. Biochemical Journal **368**, 377-396

3 Lorand, L. and Graham, R. M. (2003) Transglutaminases: crosslinking enzymes with pleiotropic functions. Nature Reviews Molecular Cell Biology **4**, 140-156

4 Folk, J. E., Cole, P. W. and Mullooly, J. P. (1968) Mechanim of action of guinea pig liver transglutaminase. V. The hydrolysis reaction. Journal of Biological Chemistry **243**, 418-427

5 Mehta, K. (2005) Mammalian transglutaminases: a family portrait. Progress in Experimental Tumor Research **38**, 1-18

6 Wada, F., Nakamura, A., Masutani, T., Ikura, K., Maki, M. and Hitomi, K. (2002) Identification of mammalian-type transglutaminase in Physarum polycephalum. Evidence from the cDNA sequence and involvement of GTP in the regulation of transamidating activity. European Journal of Biochemistry **269**, 3451-3460

7 Aeschlimann, D. and Paulsson, M. (1994) Transglutaminases: protein cross-linking enzymes in tissues and body fluids. Thrombosis & Haemostasis **71**, 402-415

8 Folk, J. E. and Chung, S. I. (1985) Transglutaminases. Methods in Enzymology **113**, 358-375

9 Clarke, D. D., Mycek, M.J., Neidle, A. and Waelsch, H. (1959) The Incorporation of Amines into Proteins. Archives of Biochemistry And Biophysics **79**, 338-354

10 de Macédo, P., Marrano, C. and Keillor, J. W. (2000) A direct continuous spectrophotometric assay for transglutaminase activity. Analytical Biochemistry **285**, 16-20

11 Leblanc, A., Gravel, C., Labelle, J. and Keillor, J. W. (2001) Kinetic studies of guinea pig liver transglutaminase reveal a general-base-catalyzed deacylation mechanism. Biochemistry **40**, 8335-8342

12 Halim, D., Caron, K. and Keillor, J. W. (2007) Synthesis and evaluation of peptidic maleimides as transglutaminase inhibitors. Bioorg Med Chem Lett **17,** 305-8

13 Gillet, S. M., Pelletier, J. N. and Keillor, J. W. (2005) A direct fluorometric assay for tissue transglutaminase. Analytical Biochemistry **347**, 221-226

14 Choi, K., Siegel, M., Piper, J. L., Yuan, L., Cho, E., Strnad, P., Omary, B., Rich, K. M. and Khosla, C. (2005) Chemistry and biology of dihydroisoxazole derivatives: selective inhibitors of human transglutaminase 2. Chemistry & Biology **12**, 469-475

15 Pardin, C., Gillet, S. M. and Keillor, J. W. (2006) Synthesis and evaluation of peptidic irreversible inhibitors of tissue transglutaminase. Bioorg Med Chem **14**, 8379-8385

16 Gillet, S. M., Chica, R. A., Keillor, J. W. and Pelletier, J. N. (2004) Expression and rapid purification of highly active hexahistidine-tagged guinea pig liver transglutaminase. Protein Expression & Purification **33**, 256-264

17 Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene **77**, 51-59

18 Chica, R. A., Gillet, S. M., Dryburgh, F.-J., Keillor, J. W. and Pelletier, J. N. (2006) Homology modeling of guinea pig liver transglutaminase and mutagenesis of conserved active-site residues Tyr519 and Cys336. *submitted.*

19 Yokoyama, K., Kikuchi, Y. and Yasueda, H. (1998) Overproduction of DnaJ in *Escherichia coli* improves in vivo solubility of the recombinant fish-derived transglutaminase. Bioscience, Biotechnology & Biochemistry **62**, 1205-1210

20 Pelletier, J. N. and Denault, M. G. (2007) Protein library design and screening: Working out the probabilities. *In* Methods Mol. Biol.: Protein Engineering Protocols, Humana Press

21 Chica, R. A., Gagnon, P., Keillor, J. W. and Pelletier, J. N. (2004) Tissue transglutaminase acylation: Proposed role of conserved active-site Tyr and Trp residues revealed by molecular modeling of peptide substrate binding. Protein Science **13**, 979-991

22 Case, A. and Stein, R. L. (2003) Kinetic analysis of the action of tissue transglutaminase on peptide and protein substrates. Biochemistry **42**, 9466-9481

23 Nemes, Z., Petrovski, G., Csosz, E. and Fesus, L. (2005) Structure-function relationships of transglutaminases--a contemporary view. Progress in Experimental Tumor Research **38**, 19-36

24 Kashiwagi, T., Yokoyama, K., Ishikawa, K., Ono, K., Ejima, D., Matsui, H. and Suzuki, E. (2002) Crystal structure of microbial transglutaminase from *Streptoverticillium mobaraense*. Journal of Biological Chemistry **277**, 44252-44260

25 Lombard, C., Saulnier, J. and Wallach, J. M. (2005) Recent trends in protease-catalyzed peptide synthesis. Protein & Peptide Letters **12**, 621-629

26 Kumar, D. and Bhalla, T. C. (2005) Microbial proteases in peptide synthesis: approaches and applications. Appl. Microbiol. Biotechnol. **68**, 726-736

27 Riechmann, L. and Kasche, V. (1985) Peptide synthesis catalyzed by the serine proteinases chymotrypsin and trypsin. Biochimica & Biophysica Acta **830**, 164-172

28 Kullmann, W. (1984) Kinetics of chymotrypsin- and papain-catalysed synthesis of [leucine]enkephalin and [methionine]enkephalin. Biochemical Journal **220**, 405-416

29 Kullmann, W. (1980) Proteases as catalysts for enzymic syntheses of opioid peptides. Journal of Biological Chemistry **255**, 8234-8238

30 Murthy, S. N., Iismaa, S., Begg, G., Freymann, D. M., Graham, R. M. and Lorand, L. (2002) Conserved tryptophan in the core domain of transglutaminase is essential for catalytic activity. Proceedings of the National Academy of Sciences of the United States of America **99**, 2738-2742

31 Iismaa, S. E., Holman, S., Wouters, M. A., Lorand, L., Graham, R. M. and Husain, A. (2003) Evolutionary specialization of a tryptophan indole group for transition-state stabilization by eukaryotic transglutaminases. Proceedings of the National Academy of Sciences of the United States of America **100**, 12636-12641

32 Pedersen, L. C., Yee, V. C., Bishop, P. D., Le Trong, I., Teller, D. C. and Stenkamp, R. E. (1994) Transglutaminase factor XIII uses proteinase-like catalytic triad to crosslink macromolecules. Protein Science **3**, 1131-1135

33 Yee, V. C., Pedersen, L. C., Le Trong, I., Bishop, P. D., Stenkamp, R. E. and Teller, D. C. (1994) Three-dimensional structure of a transglutaminase: human blood coagulation factor XIII. Proceedings of the National Academy of Sciences of the United States of America **91**, 7296-7300

CHAPITRE 7

Conclusion

7.1 Conclusion

Le but de notre projet de recherche était de développer de nouveaux biocatalyseurs pour la synthèse des peptides. Pour atteindre cet objectif général, nous avons procédé à la modification de la spécificité de la TGase de foie de cobaye par évolution dirigée selon une approche semi-aléatoire. Cette approche a été choisie, car elle exploite les données connues sur la structure et/ou la fonction de certains résidus, générant ainsi des banques de mutants « intelligentes » ayant une plus grande probabilité de contenir des mutants aux propriétés désirées. Ceci a l'avantage de ne pas requérir une méthode de criblage à haut débit, puisque la taille des banques de mutants est limitée selons nos besoins. L'approche semi-aléatoire semblait donc être le meilleur choix pour atteindre notre objectif.

Afin d'appliquer cette approche, nous avions besoin d'information structurelle et fonctionnelle qui n'était pas disponible dans la littérature. Nous avons obtenu plus d'information structurelle par l'identification des résidus du site actif de la TGase tissulaire impliqués dans la liaison des substrats. Pour identifier ces résidus, nous avons utilisé une approche combinant la cinétique enzymatique et la modélisation moléculaire qui nous a permis de développer un modèle moléculaire de la TGase tissulaire avec le substrat donneur d'acyle Cbz-Gln-Gly lié au site actif. Ce modèle nous a permis de proposer une séquence d'événements lors de l'étape d'acylation du mécanisme des TGases. De plus, il nous a permis d'identifier les résidus du site actif qui sont en contact direct avec le substrat.

Par la suite, nous avons conçu une méthode efficace et rapide d'expression et de purification de la TGase de foie de cobaye chez *E. coli* de sorte à pouvoir obtenir l'enzyme nécessaire pour le développement des méthodes expérimentales essentielles au projet. Ce protocole nous a permis d'obtenir de l'enzyme hautement pure ayant l'activité spécifique la plus élevée rapportée dans la littérature. De plus, l'expression en présence des chaperones DnaK et DnaJ nous a permis d'obtenir une grande quantité d'enzyme recombinante ayant des paramètres cinétiques similaires à ceux de l'enzyme native purifiée à partir d'organes

animaux. La conception de cette méthode présente aussi l'avantage de permettre les modifications génétiques de l'enzyme recombinante.

Toujours dans le but d'obtenir de l'information structurelle et fonctionnelle requise pour développer des banques « intelligentes », nous avons généré le premier modèle par homologie de la TGase de foie de cobaye puisque aucune structure cristalline n'avait été résolue pour cette enzyme. Ce modèle a démontré que la structure de la TGase de foie de cobaye est très similaire à celle de la TGase tissulaire de *Pagrus major* que nous avons utilisée pour développer le modèle de liaison du substrat donneur Cbz-Gln-Gly. Le modèle représente un nouvel outil pour toute étude structurelle de l'enzyme de foie de cobaye pour laquelle aucune structure cristalline n'a été résolue. Nous avons aussi procédé à la mutagenèse des résidus Tyr519 et Cys336 du site actif de la TGase de foie de cobaye, seuls ou en combinaison, et avons démontré la tolérance aux mutations du site actif. Le modèle et les résultats de mutagenèse ont validé l'utilisation de la TGase de foie de cobaye pour l'évolution dirigée par une approche semi-aléatoire.

Enfin, ces avancées nous ont permis de développer une méthode de criblage fluorimétrique basée sur des substrats donneurs contenant un ester 7-hydroxycoumaryl permettant de détecter le changement de spécificité de la TGase. Nous avons découvert que la TGase de foie de cobaye de type sauvage catalysait déjà la formation de liaisons peptidiques entre un dérivé ester 7-hydroxycoumaryl de la Gly et des dérivés de petits acides aminés polaires ou hydrophobes grâce à une activité peptide synthase intrinsèque. La spécificité de l'activité peptide synthase de la TGase de foie de cobaye a été modifiée par évolution dirigée selon une approche semi-aléatoire. En exploitant l'information structurelle obtenue par modélisation moléculaire, nous avons procédé à la création de cinq banques de mutants ciblant certains résidus du site actif. Plus de 3000 clones provenant de ces cinq banques ont été criblés, ce qui nous a permis d'identifier un mutant (Trp332Phe) pouvant former des liens peptidiques entre la Gly et des dérivés de L-Phe et de L-Tyr, ce qui n'est pas le cas de l'enzyme de type sauvage. Ce mutant a une efficacité catalytique

similaire à celle de la papaïne pour la synthèse du lien peptidique Gly-Phe et ne devrait pas hydrolyser le produit, ce qui démontre qu'il pourrait être avantageux comme biocatalyseur pour la synthèse enzymatique de peptides.

En conclusion, nous avons démontré l'activité peptide synthase intrinsèque de la TGase de foie de cobaye de type sauvage et avons modifié sa spécificité par évolution dirigée selon une approche semi-aléatoire. Bien que sa spécificité demeure étroite, cette enzyme représente un nouvel outil pour la synthèse enzymatique des peptides et sa spécificité pourrait complémenter celle des protéases utilisées pour la biocatalyse de ce type de réactions.

7.2 Perspectives

Ce projet a permis de démontrer que la TGase de foie de cobaye représente un outil prometteur pour la synthèse enzymatique des peptides. Afin qu'elle soit pleinement exploitée, une caractérisation plus approfondie est en cours. Cette caractérisation inclut l'analyse cinétique avec une plus grande variété de substrats accepteurs potentiels de sorte à clarifier les limites de la spécificité de la TGase et la mesure du rendement de la réaction de synthèse peptidique catalysée par l'enzyme de sorte à connaître son efficacité.

À plus long terme, plusieurs propriétés de la TGase pourraient être améliorées ou modifiées pour qu'elle devienne un meilleur catalyseur de la synthèse des peptides. Des changements supplémentaires à la cavité du site actif pourraient être envisagées pour que la spécificité envers les substrats donneurs et accepteurs soit modifiée. Pour ce faire, les cinq banques pourraient être combinées entre elles et de toutes nouvelles banques de mutants ciblant d'autres résidus pourraient être créées. Aussi, les cinq banques pourraient être criblées avec de nouveaux substrats donneurs ou accepteurs potentiels afin que de nouvelles spécificités soient identifiées. Une méthode de criblage ayant une limite de détection plus basse et étant basée sur le principe du FRET (*fluorescence resonance energy transfer*)

pourrait être conçue. Une méthode de ce genre permettrait de détecter la formation directe de liens peptidiques plutôt que la libération d'un fluorophore à l'étape d'acylation, comme c'est le cas avec la méthode présentée au chapitre 6. Finalement, l'enzyme pourrait être soumise à une évolution dirigée selon une approche aléatoire afin que son efficacité catalytique ou sa stabilité soit augmentée.

Bibliographie

1 Ege, S. N. (1994) Organic chemistry. D.C. Heath and Company, Toronto

2 Anastas, P. T. and Williamson, T. C. (1996) Green chemistry: An overview. In Green chemistry: Designing chemistry for the environment (Anastas, P. T. and Williamson, T. C., eds.), pp. 1-17, American Chemical Society, Washington D.C.

3 Alcalde, M., Ferrer, M., Plou, F. J. and Ballesteros, A. (2006) Environmental biocatalysis: from remediation with enzymes to novel green processes. Trends in Biotechnology **24**, 281-287

4 Anastas, P. T. and Warner, J. C. (1998) Green chemistry. Theory and practice. Oxford University Press, New York

5 Wescott, C. R. and Klibanov, A. M. (1994) The solvent dependence of enzyme specificity. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology **1206**, 1-9

6 Zhao, H., Chockalingam, K. and Chen, Z. (2002) Directed evolution of enzymes and pathways for industrial biocatalysis. Current Opinion in Biotechnology **13**, 104-110

7 Panke, S., Held, M. and Wubbolts, M. (2004) Trends and innovations in industrial biocatalysis for the production of fine chemicals. Current Opinion in Biotechnology **15**, 272-279

8 Zhang, H., Xu, X., Mu, H., Nilsson, J., Adler-Nissen, J. and Høy, C. (2000) Lipozyme IM-catalyzed interesterification for the production of margarine fats in a 1 kg scale stirred tank reactor. European Journal of Lipid Science & Technology **102**, 411-418

9 Voet, D. and Voet, J. G. (1995) Biochemistry. John Wiley & Sons, Inc., Toronto

10 Erbeldinger, M., Ni, X. and Halling, P. J. (2001) Kinetics of enzymatic solid-to-solid peptide synthesis: synthesis of Z-aspartame and control of acid-base conditions by using inorganic salts. Biotechnology & Bioengineering **72**, 69-76

11 Andersson, L., Blomberg, L., Flegel, M., Lepsa, L., Nilsson, B. and Verlander, M. (2000) Large-scale synthesis of peptides. Biopolymers **55**, 227-250

12 Mayo, K. H. (2000) Recent advances in the design and construction of synthetic peptides: for the love of basics or just for the technology of it. Trends in Biotechnology **18**, 212-217

13 Jenssen, H., Hamill, P. and Hancock, R. E. (2006) Peptide antimicrobial agents. Clinical Microbiology Reviews **19**, 491-511

14 Kumar, D. and Bhalla, T. C. (2005) Microbial proteases in peptide synthesis: approaches and applications. Applied Microbiology & Biotechnology **68**, 726-736

15 Carpino, L. A. (1993) 1-Hydroxy-7-azabenzotriazole. An Efficient Peptide Coupling Additive. Journal of the American Chemical Society **115**, 4397-4398

16 Knorr, R., Trzeciak, A., Bannwarth, W. and Gillessen, D. (1989) New coupling reagents in peptide chemistry. Tetrahedron Letters **30**, 1927-1930

17 Coste, J., Le-Nguyen, D. and Castro, B. (1990) PyBOP(R): A new peptide coupling reagent devoid of toxic by-product. Tetrahedron Letters **31**, 205-208

18 Kullmann, W. (1987) Enzymatic Peptide Synthesis. CRC Press, Inc., Boca Raton, Florida

19 Lombard, C., Saulnier, J. and Wallach, J. M. (2005) Recent trends in protease-catalyzed peptide synthesis. Protein & Peptide Letters **12**, 621-629

20 Griffin, M., Casadio, R. and Bergamini, C. M. (2002) Transglutaminases: nature's biological glues. Biochemical Journal **368**, 377-396

21 Autuori, F., Farrace, M. G., Oliverio, S., Piredda, L. and Piacentini, M. (1998) "Tissue" transglutaminase and apoptosis. Advances in Biochemical Engineering & Biotechnology **62**, 129-136

22 Abe, S., Yamashita, K., Kohno, H. and Ohkubo, Y. (2000) Involvement of transglutaminase in the receptor-mediated endocytosis of mouse peritoneal macrophages. Biological and Pharmaceutical Bulletin **23**, 1511-1513

23 Chen, J. S. and Mehta, K. (1999) Tissue transglutaminase: an enzyme with a split personality. International Journal of Biochemistry & Cell Biology **31**, 817-836

24 Shridas, P., Sharma, Y. and Balasubramanian, D. (2001) Transglutaminase-mediated cross-linking of alpha-crystallin: structural and functional consequences. FEBS Letters **499**, 245-250

25 Shan, L., Molberg, Ø., Parrot, I., Hausch, F., Filiz, F., Gray, G. M., Sollid, L. M. and Khosla, C. (2002) Structural basis for gluten intolerance in celiac sprue. Science **297**, 2275-2279

26 Schroeder, W. T., Thacher, S. M., Stewart-Galetka, S., Annarella, M., Chema, D., Siciliano, M. J., Davies, P. J., Tang, H. Y., Sowa, B. A. and Duvic, M. (1992) Type I keratinocyte transglutaminase: expression in human skin and psoriasis. Journal of Investigative Dermatology **99**, 27-34

27 Lorand, L. (1996) Neurodegenerative diseases and transglutaminase. Proceedings of the National Academy of Science of the United States of America **93**, 14310-14313

28 Folk, J. E. (1969) Mechanism of action of guinea pig liver transglutaminase. VI. Order of substrate addition. Journal of Biological Chemistry **244**, 3707-3713

29 Ahvazi, B. and Steinert, P. M. (2003) A model for the reaction mechanism of the transglutaminase 3 enzyme. Experimental & Molecular Medicine **35**, 228-242

30 Yee, V. C., Pedersen, L. C., Le Trong, I., Bishop, P. D., Stenkamp, R. E. and Teller, D. C. (1994) Three-dimensional structure of a transglutaminase: human blood coagulation factor XIII. Proceedings of the National Academy of Sciences of the United States of America **91**, 7296-7300

31 Case, A. and Stein, R. L. (2003) Kinetic analysis of the action of tissue transglutaminase on peptide and protein substrates. Biochemistry **42**, 9466-9481

32 Leblanc, A., Gravel, C., Labelle, J. and Keillor, J. W. (2001) Kinetic studies of guinea pig liver transglutaminase reveal a general-base-catalyzed deacylation mechanism. Biochemistry **40**, 8335-8342

33 Pasternack, R., Dorsch, S., Otterbach, J. T., Robenek, I. R., Wolf, S. and Fuchsbauer, H. L. (1998) Bacterial pro-transglutaminase from Streptoverticillium mobaraense--purification, characterisation and sequence of the zymogen. European Journal of Biochemistry **257**, 570-576

34 Della Mea, M., Caparros-Ruiz, D., Claparols, I., Serafini-Fracassini, D. and Rigau, J. (2004) AtPng1p. The first plant transglutaminase. Plant Physiology **135**, 2046-2054

35 Iranzo, M., Aguado, C., Pallotti, C., Canizares, J. V. and Mormeneo, S. (2002) Transglutaminase activity is involved in Saccharomyces cerevisiae wall construction. Microbiology **148**, 1329-1334

36 Madi, A., Punyiczki, M., di Rao, M., Piacentini, M. and Fesus, L. (1998) Biochemical characterization and localization of transglutaminase in wild-type and cell-death mutants of the nematode Caenorhabditis elegans. European Journal of Biochemistry **253**, 583-590

37 Mehta, K. (2005) Mammalian transglutaminases: a family portrait. Progress in Experimental Tumor Research **38**, 1-18

38 Cooper, A. J. L., Jeitner, T. M. and Blass, J. P. (2002) The role of transglutaminases in neurodegenerative diseases: overview. Neurochemistry International **40**, 1-5

39 Lorand, L. and Graham, R. M. (2003) Transglutaminases: crosslinking enzymes with pleiotropic functions. Nature Reviews Molecular Cell Biology **4**, 140-156

40 Lai, T. S., Slaughter, T. F., Peoples, K. A. and Greenberg, C. S. (1999) Site-directed mutagenesis of the calcium-binding site of blood coagulation factor XIIIa. Journal of Biological Chemistry **274**, 24953-24958

41 Steinert, P. M., Kim, S. Y., Chung, S. I. and Marekov, L. N. (1996) The transglutaminase 1 enzyme is variably acylated by myristate and palmitate during differentiation in epidermal keratinocytes. Journal of Biological Chemistry **271**, 26242-26250

42 Boeshans, K. M., Mueser, T. C. and Ahvazi, B. (2007) A three-dimensional model of the human transglutaminase 1: insights into the understanding of lamellar ichthyosis. Journal of Molecular Modeling **13**, 233-246

43 Lesort, M., Attanavanich, K., Zhang, J. and Johnson, G. V. (1998) Distinct nuclear localization and activity of tissue transglutaminase. Journal of Biological Chemistry **273**, 11991-11994

44 Nakaoka, H., Perez, D. M., Baek, K. J., Das, T., Husain, A., Misono, K., Im, M. J. and Graham, R. M. (1994) Gh: a GTP-binding protein with transglutaminase activity and receptor signaling function. Science **264**, 1593-1596

45 Hasegawa, G., Suwa, M., Ichikawa, Y., Ohtsuka, T., Kumagai, S., Kikuchi, M., Sato, Y. and Saito, Y. (2003) A novel function of tissue-type transglutaminase: protein disulphide isomerase. Biochemical Journal **373**, 793-803

46 Mishra, S. and Murphy, L. J. (2004) Tissue transglutaminase has intrinsic kinase activity: identification of transglutaminase 2 as an insulin-like growth factor-binding protein-3 kinase. Journal of Biological Chemistry **279**, 23863-23868

47 Fesus, L. and Piacentini, M. (2002) Transglutaminase 2: an enigmatic enzyme with diverse functions. Trends in Biochemical Sciences **27**, 534-539

48 Fesus, L., Thomazy, V. and Falus, A. (1987) Induction and activation of tissue transglutaminase during programmed cell death. FEBS Letters **224**, 104-108

49 Ahvazi, B., Boeshans, K. M. and Steinert, P. M. (2004) Crystal structure of transglutaminase 3 in complex with GMP: structural basis for nucleotide specificity. Journal of Biological Chemistry **279**, 26716-26725

50 Candi, E., Paradisi, A., Terrinoni, A., Pietroni, V., Oddi, S., Cadot, B., Jogini, V., Meiyappan, M., Clardy, J., Finazzi-Agro, A. and Melino, G. (2004) Transglutaminase 5 is regulated by guanine-adenine nucleotides. Biochemical Journal **381**, 313-319

51 Kashiwagi, T., Yokoyama, K., Ishikawa, K., Ono, K., Ejima, D., Matsui, H. and Suzuki, E. (2002) Crystal structure of microbial transglutaminase from Streptoverticillium mobaraense. Journal of Biological Chemistry **277**, 44252-44260

52 Fox, B. A., Yee, V. C., Pedersen, L. C., Le Trong, I., Bishop, P. D., Stenkamp, R. E. and Teller, D. C. (1999) Identification of the calcium binding site and a novel ytterbium site in blood coagulation factor XIII by x-ray crystallography. Journal of Biological Chemistry **274**, 4917-4923

53 Pedersen, L. C., Yee, V. C., Bishop, P. D., Le Trong, I., Teller, D. C. and Stenkamp, R. E. (1994) Transglutaminase factor XIII uses proteinase-like catalytic triad to crosslink macromolecules. Protein Science **3**, 1131-1135

54 Weiss, M. S., Metzner, H. J. and Hilgenfeld, R. (1998) Two non-proline cis peptide bonds may be important for factor XIII function. FEBS Letters **423**, 291-296

55 Liu, S., Cerione, R. A. and Clardy, J. (2002) Structural basis for the guanine nucleotide-binding activity of tissue transglutaminase and its regulation of transamidation activity. Proceedings of the National Academy of Sciences of the United States of America **99**, 2743-2747

56 Noguchi, K., Ishikawa, K., Yokoyama, K., Ohtsuka, T., Nio, N. and Suzuki, E. (2001) Crystal structure of red sea bream transglutaminase. Journal of Biological Chemistry **276**, 12055-12059

57 Ahvazi, B., Kim, H. C., Kee, S. H., Nemes, Z. and Steinert, P. M. (2002) Three-dimensional structure of the human transglutaminase 3 enzyme: binding of calcium ions changes structure for activation. EMBO Journal **21**, 2055-2067

58 Ahvazi, B., Boeshans, K. M., Idler, W., Baxa, U. and Steinert, P. M. (2003) Roles of calcium ions in the activation and activity of the transglutaminase 3 enzyme. Journal of Biological Chemistry **278**, 23834-23841

59 Ahvazi, B., Boeshans, K. M., Idler, W., Baxa, U., Steinert, P. M. and Rastinejad, F. (2004) Structural basis for the coordinated regulation of transglutaminase 3 by guanine nucleotides and calcium/magnesium. Journal of Biological Chemistry **279**, 7180-7192

60 Lai, T. S., Slaughter, T. F., Koropchak, C. M., Haroon, Z. A. and Greenberg, C. S. (1996) C-terminal deletion of human tissue transglutaminase enhances magnesium-dependent GTP/ATPase activity. Journal of Biological Chemistry **271**, 31191-31195

61 Lai, T. S., Achyuthan, K. E., Santiago, M. A. and Greenberg, G. S. (1994) Carboxyl-terminal truncation of recombinant factor XIII A-chains. Characterization of minimum structural requirement for transglutaminase activity. Journal of Biological Chemistry **269**, 24596-24601

62 Wilming, M., Iffland, A., Tafelmeyer, P., Arrivoli, C., Saudan, C. and Johnsson, K. (2002) Examining reactivity and specificity of cytochrome c peroxidase by using combinatorial mutagenesis. Chembiochem **3**, 1097-1104

63 Facchiano, F. and Facchiano, A. (2005) Transglutaminases and their substrates. Progress in Experimental Tumor Research **38**, 37-57

64 Choi, K., Siegel, M., Piper, J. L., Yuan, L., Cho, E., Strnad, P., Omary, B., Rich, K. M. and Khosla, C. (2005) Chemistry and biology of dihydroisoxazole derivatives: selective inhibitors of human transglutaminase 2. Chemistry & Biology **12**, 469-475

65 Hausch, F., Halttunen, T., Maki, M. and Khosla, C. (2003) Design, synthesis, and evaluation of gluten peptide analogs as selective inhibitors of human tissue transglutaminase. Chemistry & Biology **10**, 225-231

66 Clarke, D. D., Mycek, M.J., Neidle, A. and Waelsch, H. (1959) The Incorporation of Amines into Proteins. Archives of Biochemistry and Biophysics **79**, 338-354

67 Aeschlimann, D. and Paulsson, M. (1994) Transglutaminases: protein cross-linking enzymes in tissues and body fluids. Thrombosis & Haemostasis **71**, 402-415

68 Folk, J. E. and Chung, S. I. (1985) Transglutaminases. Methods in Enzymology **113**, 358-375

69 de Macedo, P., Marrano, C. and Keillor, J. W. (2000) A direct continuous spectrophotometric assay for transglutaminase activity. Analytical Biochemistry **285**, 16-20

70 Chica, R. A., Gagnon, P., Keillor, J. W. and Pelletier, J. N. (2004) Tissue transglutaminase acylation: Proposed role of conserved active site Tyr and Trp residues revealed by molecular modeling of peptide substrate binding. Protein Science **13**, 979-991

71 Gillet, S. M., Chica, R. A., Keillor, J. W. and Pelletier, J. N. (2004) Expression and rapid purification of highly active hexahistidine-tagged guinea pig liver transglutaminase. Protein Expression & Purification **33**, 256-264

72 Day, N. and Keillor, J. W. (1999) A continuous spectrophotometric linked enzyme assay for transglutaminase activity. Analytical Biochemistry **274**, 141-144

73 Halim, D., Caron, K. and Keillor, J. W. (2006) Synthesis and evaluation of peptidic maleimides as transglutaminase inhibitors. Bioorganic & Medicinal Chemistry Letters

74 Dalby, P. A. (2003) Optimising enzyme function by directed evolution. Current Opinion in Structural Biology **13**, 500-505

75 Jaeger, K. E. and Eggert, T. (2004) Enantioselective biocatalysis optimized by directed evolution. Current Opinion in Biotechnology **15**, 305-313

76 Jestin, J. L. and Kaminski, P. A. (2004) Directed enzyme evolution and selections for catalysis based on product formation. Journal of Biotechnology **113**, 85-103

77 Tao, H. and Cornish, V. W. (2002) Milestones in directed enzyme evolution. Current Opinion in Chemical Biology **6**, 858-864

78 Williams, G. J., Nelson, A. S. and Berry, A. (2004) Directed evolution of enzymes for biocatalysis and the life sciences. Cellular and Molecular Life Sciences **61**, 3034-3046

79 Stemmer, W. P. (1994) Rapid evolution of a protein in vitro by DNA shuffling. Nature **370**, 389-391

80 Stemmer, W. P. (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proceedings of the National Academy of Sciences of the United States of America **91**, 10747-10751

81 Koga, Y., Kato, K., Nakano, H. and Yamane, T. (2003) Inverting enantioselectivity of Burkholderia cepacia KWI-56 lipase by combinatorial mutation and high-throughput screening using single-molecule PCR and in vitro expression. Journal of Molecular Biology **331**, 585-592