

**STRUCTURAL DETERMINANTS IN THE FOLDING OF
EPIDERMAL GROWTH FACTOR (EGF)-LIKE DOMAINS**

NG AH SOCK ANGIE
(B.Sc.(Hons.), NTU)

**A THESIS SUBMITTED
FOR THE DEGREE OF MASTER OF SCIENCE
DEPARTMENT OF BIOLOGICAL SCIENCES
NATIONAL UNIVERSITY OF SINGAPORE**

2011

Acknowledgements

I owe my deepest gratitude to my supervisor, Professor R.M. Kini, for his guidance throughout the course of my research project. I have learnt a lot from him — Not just in the field of protein chemistry, but also useful skills in research such as time management and effective planning. I would also like to thank him for his constant encouragement over the past two years — It had truly motivated me to pursue a career in scientific research. I am now looking forward to the possibility of enrolling in a PhD program to gain more professional training as a researcher.

I am also heartily thankful to people who had helped me in one way or another in my research work: Dr. Koh Cho Yeow for his patient guidance through the initial phase of my research project; Assistant Professor Kim Chu-Young and Mr. Sathya Dev Unudurthi for their help in imparting the skill of manual solid phase peptide synthesis to me; Mr. Vallerinteavide Mavelli Girish for his constant help in troubleshooting the ÄKTA purifier system and the Perkin Elmer ESI-MS system; Miss Tay Bee Ling for her prompt attendance to matters regarding product purchases and logistical issues.

Not forgetting all members of the Protein Science Laboratory — Thank you very much for your kind support! I enjoy talking to all of you, sharing ideas and aspirations. You are a fun-loving group of people, creating a positive atmosphere despite the stressful demands of our every day life. These had made my research experience in the lab a truly unforgettable one.

Ng Ah Sock Angie
2011

Table of Contents

Acknowledgements	I
Table of Contents	II
Summary	VI
List of Tables	VII
List of Figures	VIII
List of Abbreviations	XII
Chapter 1: Introduction	1
1.1 The Protein Folding Problem	2
1.1.1 The folding code	2
1.1.2 The folding pathway	8
1.2 Disulfide Bonds as Probes of Protein Folding	12
1.2.1 Trapped disulfide-containing intermediates for the study of protein folding pathway	12
1.2.2 Disulfide-connectivity based structural isoforms for the study of protein folding code	13
1.3 The Canonical Fold of the EGF-like domain	17
1.3.1 Description of the canonical EGF-like domain fold	17
1.3.2 Significance of studying the protein folding code of EGF-like domain	19
1.4 Thrombomodulin and Its Role in the Anti-coagulation Pathway	22
1.4.1 TM as a regulator of the coagulation cascade	22
1.4.2 Structure-function relationship of TM: Role of the fourth to sixth EGF-like domains	24
1.5 Thrombomodulin EGF-like Domain 4 and 5:	28
Models in the Study of the EGF-like Domain Folding Code	28

1.5.1	TM EGF D4 <i>versus</i> TM EGF D5: Canonical <i>versus</i> non-canonical EGF-like domain fold	28
1.5.2	TM EGF D4 and TM EGF D5 as models to identify the structural determinants of the canonical EGF-like domain fold	32
1.6	Objectives and Scope of the Thesis	34
Chapter 2: Materials and Methods		35
2.1	Peptide Synthesis and Purification	36
2.1.1	Peptide synthesis	36
2.1.2	Peptide cleavage, deprotection and isolation	37
2.1.3	Peptide purification	38
2.1.4	Electrospray ionization-mass spectrometry (ESI-MS)	38
2.2	Regioselective Synthesis of Structural Isoforms	39
2.2.1	Formation of the first disulfide bridge	39
2.2.2	Formation of the second disulfide bridge: Iodine mediated simultaneous deprotection/oxidation	40
2.3	Oxidative Folding of Fully Reduced Peptides	41
2.3.1	Air oxidation	41
2.3.2	Oxidation in the presence of redox reagents	41
2.4	Chromatographic Separation of Structural Iso-forms Obtained from Oxidative Folding Studies	42
2.4.1	Structural isoforms of t-TM EGF D4	42
2.4.2	Structural isoforms of t-TM EGF D5	42
2.4.3	Structural isoforms of t-TM EGF D4 (Y25T)	42
2.4.4	Calculation of peak area	43
2.4.5	Statistical analysis	43
Chapter 3: Results and Discussion		46
3.1	Synthesis of Truncated TM EGF D4 and TM EGF D5 Structural Isoforms	47
3.1.1	Elution characteristics of t-TM EGF D4 structural isoforms	48
3.1.2	Elution characteristics of t-TM EGF D5 structural isoforms	50
3.2	The <i>in vitro</i> Folding Tendencies of t-TM EGF D4 and t-TM EGF D5	52
3.2.1	<i>In vitro</i> oxidative folding of t-TM EGF D4	52

3.2.2	<i>In vitro</i> oxidative folding of t-TM EGF D5	57
3.2.3	Truncated TM EGF D4 and TM EGF D5 preferentially fold into their respective native isoform	60
3.3	Contribution of Side-chain Interactions in the Folding Tendencies of t-TM EGF D4 and t-TM EGF D5	62
3.3.1	<i>In vitro</i> oxidative folding of t-TM EGF D4 in the presence of 6 M Gn.HCl	62
3.3.2	<i>In vitro</i> oxidative folding of t-TM EGF D5 in the presence of 6 M Gn.HCl	66
3.3.3	Side-chain interaction is necessary for the canonical C ₁ -C ₃ , C ₂ -C ₄ fold of the EGF-like domain	67
3.4	Contribution of Hydrophobic Interactions in the Folding Tendencies of t-TM EGF D4 and t-TM EGF D5	80
3.4.1	<i>In vitro</i> oxidative folding of t-TM EGF D4 in the presence of 0.5 M NaCl	80
3.4.2	<i>In vitro</i> oxidative folding of t-TM EGF D5 in the presence of 0.5 M NaCl	80
3.4.3	Hydrophobic interaction is necessary for the canonical C ₁ -C ₃ , C ₂ -C ₄ fold of the EGF-like domain	81
3.5	Identification of Key Hydrophobic Residues as Structural Determinants of the Canonical EGF-like Domain fold in t-TM EGF D4	85
3.5.1	<i>In vitro</i> oxidative folding of t-TM EGF D4 (Y25T)	92
3.5.2	<i>In vitro</i> oxidative folding of t-TM EGF D4 (Y25T) in the presence of 6 M Gn.HCl	97
3.5.3	The hydrophobic/aromatic residue, Tyr25, as the main structural determinant of t-TM EGF D4	97

Chapter 4: Conclusion **103**

4.1	Conclusion	104
4.2	Future Work	109
4.2.1	Verifying the structural determinant of the canonical EGF-like domain fold	109
4.2.2	The role of the structural determinant in the transition state of protein folding	109
4.2.3	Extending the study to other canonical EGF-like domains	110
4.3	Implication of Findings	111

Bibliography **113**

Appendix **120**

Summary

The epidermal growth factor (EGF)-like domain is an evolutionarily conserved modular protein subunit. Despite hypervariability of amino acid sequences in their inter-cysteine region, they preferentially fold into a three-looped conformation with a disulfide pairing of C₁-C₃, C₂-C₄, C₅-C₆. To elucidate the structural determinants that dictates the canonical EGF-like domain fold, we had chosen the fourth and fifth EGF-like domain of thrombomodulin (TM) as models. While the fourth EGF-like domain folds into the canonical conformation, the fifth EGF-like domain does not and possesses an alternate disulfide pairing of C₁-C₂, C₃-C₄, C₅-C₆. We examined the folding tendencies of two synthetic peptides corresponding to truncated versions of TM EGF-like domain four and five under air oxidation and redox folding conditions. By identifying the structural isoforms obtained in the folding reaction using regiospecifically-synthesized conformers as controls, we determined that the last segment of both domains (encompassing C₅ and C₆) do not influence the tendencies to fold into their respective native conformations. When folded under denaturing conditions, the folding tendency of the fourth EGF-like domain changes to that of the C₁-C₂, C₃-C₄ conformer. Conversely, the addition of denaturant did not affect the folding tendency of the fifth EGF-like domain. This suggests that side chain interactions are crucial for achieving the canonical EGF-like domain fold but not for the non-canonical fold. Folding under high salt content did not disrupt the folding tendencies of both domains and result in slight increase of the C₁-C₃, C₂-C₄ conformer in both cases. This suggests that hydrophobic interaction, but not electrostatic interaction, is the key in the achieving the canonical fold of EGF-like domains.

List of Tables

Chapter 1

Table 1.1	Effect of various TM EGF D5 structural isoforms on thrombin activity.....	31
-----------	---	----

Chapter 2

Table 2.1	Annotation for statistical formulas (Eq. 1 to Eq. 6).....	44
-----------	---	----

Chapter 3

Table 3.1	Observed versus theoretical mass of t-TM EGF D4 structural isoforms	48
Table 3.2	Observed versus theoretical mass of t-TM EGF D5 structural isoforms	50
Table 3.3	Percentages of structural isoforms obtained from oxidative folding of t-TM EGF D4 and t-TM EGF D5 in various conditions	56
Table 3.4	Percentages of structural isoforms obtained from oxidative folding of t-TM EGF D4 (Y25T) in various conditions.....	93

List of Figures

Chapter 1

Figure 1.1	<i>In vitro</i> re-folding of ribonuclease.....	4
Figure 1.2	Disulfide scaffold of α - and X-conotoxins.	7
Figure 1.3	The consensus sequence of the EGF-like domain.	18
Figure 1.4	The canonical fold of the EGF-like domain.	18
Figure 1.5	The domain organization of thrombomodulin.....	25
Figure 1.6	Ribbon model of the complex between α -thrombin and TM EGF D4-D6 [PDB: 1DX5].....	26
Figure 1.7	Solution structure of TM EGF D4 and its disulfide-connectivity.	29
Figure 1.8	Solution structure of TM EGF D5 and its disulfide-connectivity.	30

Chapter 2

Figure 2.1	Synthesis of t-TM EGF D4 and t-TM EGF D5 structural isoforms and their respective "test peptides" using Cys(Acm) and Cys(Trt).	37
------------	---	----

Chapter 3

Figure 3.1	A comparison between the disulfide-connectivity of (A) TM EGF D4 and (B) TM EGF D5.....	47
Figure 3.2	Regioselective synthesis of t-TM EGF D4 and t-TM EGF D5.	49
Figure 3.3	Analysis of t-TM EGF D4 air oxidation products by reversed-phase chromatography.....	54
Figure 3.4	Analysis of t-TM EGF D4 redox reagent-mediated oxidation products by reversed-phase chromatography.	55
Figure 3.5	Pairwise comparison of t-TM EGF D4 structural isoform proportions obtained from air oxidation and redox reagent-mediated oxidation studies.....	57
Figure 3.6	Analysis of t-TM EGF D5 air oxidation products by reversed-phase chromatography.....	58

Figure 3.7	Analysis of t-TM EGF D5 redox reagent-mediated oxidation products by reversed-phase chromatography.	59
Figure 3.8	Pairwise comparison of t-TM EGF D5 structural isoform proportions obtained from air oxidation and redox reagent-mediated oxidation studies.....	61
Figure 3.9	Analysis of t-TM EGF D4 products obtained from air oxidation in the presence of 6 M Gn.HCl by reversed-phase chromatography.	64
Figure 3.10	Analysis of t-TM EGF D4 products obtained from redox reagent-mediated oxidation in the presence of 6 M Gn.HCl by reversed-phase chromatography..	65
Figure 3.11	Pairwise comparison of t-TM EGF D4 structural isoform proportions obtained from air oxidation (with 6 M Gn.HCl) and redox reagent-mediated oxidation (with 6 M Gn.HCl) studies..	66
Figure 3.12	Analysis of t-TM EGF D5 products obtained from air oxidation in the presence of 6 M Gn.HCl by reversed-phase chromatography.	68
Figure 3.13	Analysis of t-TM EGF D5 products obtained from redox reagent-mediated oxidation in the presence of 6 M Gn.HCl by reversed-phase chromatography.....	69
Figure 3.14	Pairwise comparison of t-TM EGF D5 structural isoform proportions obtained from air oxidation (with 6 M Gn.HCl) and redox reagent-mediated oxidation (with 6 M Gn.HCl) studies..	70
Figure 3.15	Pairwise comparison of t-TM EGF D4 structural isoform proportions obtained from air oxidation and air oxidation (with 6 M Gn.HCl) studies.....	72
Figure 3.16	Pairwise comparison of t-TM EGF D4 structural isoform proportions obtained from redox reagent-mediated oxidation and redox reagent-mediated oxidation (with 6 M Gn.HCl) studies..	72
Figure 3.17	Pairwise comparison of t-TM EGF D5 structural isoform proportions obtained from air oxidation and air oxidation (with 6 M Gn.HCl) studies.....	73
Figure 3.18	Pairwise comparison of t-TM EGF D5 structural isoform proportions obtained from redox reagent-mediated oxidation and redox reagent-mediated oxidation (with 6 M Gn.HCl) studies..	74
Figure 3.19	Space-filled model of (A) t-TM EGF D4 and (B) t-TM EGF D5..	77
Figure 3.20	Analysis of (A) t-TM EGF D4 and (B) t-TM EGF D5 products obtained from redox reagent-mediated oxidation in the presence of 0.5 M NaCl by reversed-phase chromatography.	81

Figure 3.21	Pairwise comparison of t-TM EGF D4 structural isoform proportions obtained from redox reagent-mediated oxidation and redox reagent-mediated oxidation (with 0.5 M NaCl) studies.	83
Figure 3.22	Pairwise comparison of t-TM EGF D5 structural isoform proportions obtained from redox reagent-mediated oxidation and redox reagent-mediated oxidation (with 0.5 M NaCl) studies.	83
Figure 3.23	Sequence alignment of canonical EGF-like domains from various proteins.	87
Figure 3.24	Sequence alignment of t-TM EGF D4 and t-TM EGF D5 from various organisms.	87
Figure 3.25	Identification of residues that interacts with the conserved hydrophobic/aromatic residues in various canonical EGF-like domains.	89
Figure 3.26	Residues interacting with the conserved hydrophobic/aromatic residues in (A) coagulation factor VII EGF-like domain 1 and (B) Pro-neuregulin-1 EGF-like domain.	90
Figure 3.27	Residues interacting with the conserved hydrophobic/aromatic residue in the canonical EGF-like t-TM EGF D4.	91
Figure 3.28	Analysis of t-TM EGF D4 (Y25T) air oxidation products by reversed-phase chromatography.	94
Figure 3.29	Analysis of t-TM EGF D4 (Y25T) redox reagent-mediated oxidation products by reversed-phase chromatography.	95
Figure 3.30	Proportion of structural isoforms obtained from air oxidation-mediated folding of t-TM EGF D4 and t-TM EGF D4 (Y25T).	96
Figure 3.31	Proportion of structural isoforms obtained from redox reagent-mediated oxidative folding of t-TM EGF D4 and t-TM EGF D4 (Y25T).	96
Figure 3.32	Analysis of t-TM EGF D4 (Y25T) products obtained from air oxidation in the presence of 6 M Gn.HCl by reversed-phase chromatography.	98
Figure 3.33	Analysis of t-TM EGF D4 (Y25T) products obtained from redox reagent-mediated oxidation in the presence of 6 M Gn.HCl by reversed-phase chromatography.	99
Figure 3.34	Proportion of structural isoforms obtained from air oxidation-mediated folding of t-TM EGF D4 (+6 M Gn.HCl) and t-TM EGF D4 (Y25T).	100

Figure 3.35	Proportion of structural isoforms obtained from redox reagent-mediated oxidative folding of t-TM EGF D4 (+6 M Gn.HCl) and t-TM EGF D4 (Y25T).	100
Figure 3.36	Comparison of structural isoform proportions obtained from air oxidation-mediated folding of t-TM EGF D4 (+6 M Gn.HCl), t-TM EGF D4 (Y25T) and t-TM EGF D5.	101
Figure 3.37	Comparison of structural isoform proportions obtained from redox reagent-mediated oxidative folding of t-TM EGF D4 (+6 M Gn.HCl), t-TM EGF D4 (Y25T) and t-TM EGF D5.	102

List of Abbreviations

Abbreviation	Full name
Acm	S-acetamidomethyl
ACN	Acetonitrile
AP	Appendix
BPTI	Bovine pancreatic trypsin inhibitor
CDAP	1-Cyano-4-dimethyl-aminopyridinium tetrafluoroborate
DCM	Dichloromethane
DIPEA	N,N-diisopropyl-ethylamine
DMF	N,N-dimethylformamide
DMSO	Dimethyl sulfoxide
DTNB	5,5'-dithio-bis-(2-nitrobenzoic acid)
EDT	1,2-ethanedithiol
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
ESI-MS	Electrospray ionization-mass spectrometry
FII	Factor II
FIX / FIXa	Factor IX / activated factor IX
Fmoc	9-fluorenylmethoxycarbonyl
FV / FVa	Factor V / activated factor Va
FVIII / FVIIIa	Factor VIII / activated factor VIII
FX / FXa	Factor X / activated factor Xa
FXI / FXIa	Factor XI / activated factor XI
FXIII / FXIIIa	Factor XIII / activated factor XIII
Gn.HCl	Guanidine hydrochloride
GSH	Reduced glutathione

GSSG	Oxidized glutathione
HATU	O-(7-Azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate
HFBA	Heptafluorobutyric acid
HPLC	High performance liquid chromatography
LCI	Leech carboxypeptidase inhibitor
MALDI-MS	Matrix assisted laser desorption ionization-mass spectrometry
MeOH	Methanol
NMP	N-Methyl-2-pyrrolidone
OtBu	t-butyl ester
PDB	Protein Data Bank
Pbf	2,2,4,6,7-pentamethyl-dihydrobenzofuran-5-sulfonyl
RNase	Ribonuclease
t-TM EGF D4	Truncated thrombomodulin EGF-like domain 4
t-TM EGF D5	Truncated thrombomodulin EGF-like domain 5
tBu	t-butyl ether
TCEP	tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
TM	Thrombomodulin
TM EGF D4	Thrombomodulin EGF-like domain 4
TM EGF D5	Thrombomodulin EGF-like domain 5
Trt	S-trityl



Chapter 1: Introduction



1.1 The Protein Folding Problem

Proteins obtain their native three-dimensional structure via folding from their primary structures. The question of how this folding is achieved is popularly known as the “protein folding problem”. Although protein folding can be seen as a multifaceted problem, the questions involved can be summarized into primary aspects: (a) the folding code - the mechanistic question of how the primary amino acid sequence of a protein specifies its native three-dimensional structure; (b) the folding pathway - the kinetic question regarding the route a protein take to reach its final native structure.

1.1.1 The folding code

Why do any two proteins, for example, lysozyme and ribonuclease, adopt different native three-dimensional structures? To do so, there must be a folding code that “instructs” each protein to fold into their respective native structures. What then is the nature of this folding code?

As it is through the composite of different amino acid residues that differentiates one protein from another, one would see that the folding code is embedded in the amino acid side-chains located along the polypeptide chain. These side-chains provide folding instructions in terms of various inter-atomic forces (e.g. hydrophobic interactions, Van der Waals-interactions, electrostatic interactions, hydrogen bonding) mediated by the distinct physical-chemical properties of each side-chain.

Thus, each amino acid, with its identity conferred by the nature of its side-

chain, can be perceived as a single “instructional unit” among others in the folding code. This relation is analogous to a statement in the source code of a computer program. From this viewpoint, the location of an amino acid residue in the sequence of a protein is, thus, equivalent to the logical placement of a statement among others in the source code. When a program file is being executed, an effect is produced as the computer carries out the instructions embedded in the sequence of statements of the source code. Analogously, the “execution” of the folding code will result in the folding of the polypeptide chain into its native structure based on the overall balance of inter-atomic forces dictated by the amino acid sequence.

To this end, it became apparent that the amino acid sequence in guiding protein folding is also in itself the determinant of its native three-dimensional structure. Indeed, from the famous experiments on ribonuclease (RNase), Anfinsen and colleagues demonstrated that fully reduced RNase, which lacked demonstrable secondary or tertiary structure, could spontaneously refold *in vitro* using molecular oxygen to yield a product that is indistinguishable in terms of enzymatic activity from the native enzyme [1-3] (Figure 1.1). This result leads to the postulation that a protein’s native structure is its most thermodynamically stable structure, and the information needed for the assumption of such a structure, including the correct pairing of half-cystine residues in disulfide linkages, is determined by the amino acid sequence itself. This postulation is now known as the Anfinsen’s thermodynamic hypothesis and its conclusion provides the basis of studying native structures in isolation inside a test tube rather than inside cells.

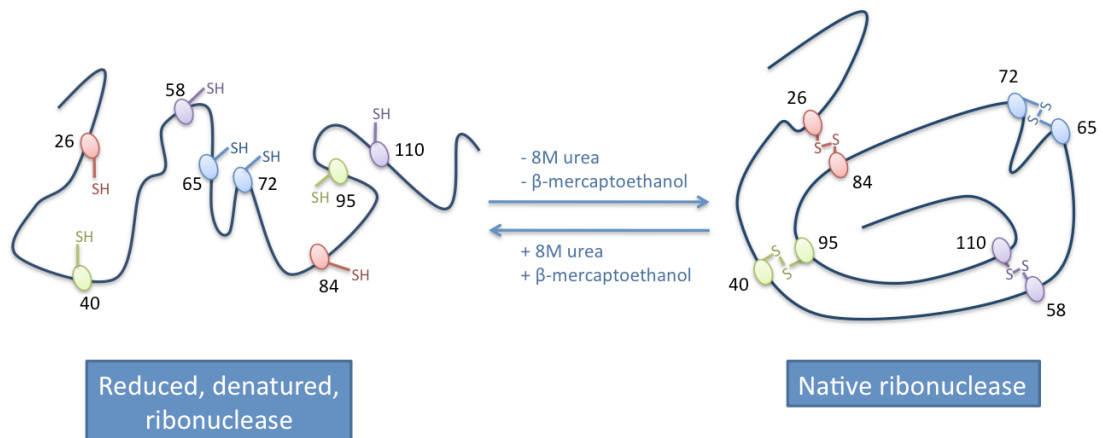


Figure 1.1 *In vitro* re-folding of ribonuclease. Reduced, denatured, ribonuclease can spontaneously refold into its native structure (with native disulfide-connectivity), upon the removal of denaturant (8M urea) and reducing agent (β -mercaptoethanol), via oxidative folding.

Although Anfinsen's thermodynamic hypothesis provides the apparent answer to the question of how proteins know *a priori* its native three-dimensional structure, the mechanistic details of how it works still remain elusive. Over the years, attempts to decipher the folding code had only lead to some general principles which are summarized below:

(a) Secondary structure propensities —

Each of the 20 natural amino acids has different intrinsic properties to populate secondary structure elements. In fact, the frequencies with which different amino acids occur in α -helices and β -sheets of natural proteins correlate with the amino acid's ability to stabilize these secondary structure elements [4]. Alanine, leucine, methionine and lysine have high propensities towards α -helices [5], whereas aromatic amino acids (tyrosine, phenylalanine and tryptophan) and β -branched amino acids (threonine, valine, isoleucine) have high propensities towards β -sheets [6]. Proline and glycine are not favored in α -helices and β -sheets and thus

have the lowest propensities for both secondary structures.

(b) Binary patterning of polar and non-polar amino acids —

Hydrophobic interaction is considered one of the dominant forces in protein folding [7]. Thus, simple binary pattern of polar and non-polar residues along the polypeptide chain has been suggested to encode low-resolution folding information which would give a protein its general topology [8]. In fact, Kamtekar *et al.* had demonstrated that *de novo*-designed binary pattern of polar and non-polar amino acid residues was sufficient to encode four-helix bundle proteins [9]. In this seminal work, combinatorial methods were used to generate a large collection of amino acid sequences where individual positions in the sequence is specified as either polar or non-polar, but the precise identity of each residue is allowed to vary. The relatively simple information encoded in the “binary code” is sufficient to generate a significant number of proteins that fold into compact α -helical structures.

(c) Complementary packing of amino acid side-chains —

If binary patterning of polar and non-polar amino acids is sufficient to specify the overall topology of the proteins, what then provide the information needed to generate the high-resolution structures of these proteins? These information come from the exact identities of the side-chains that are “complementary packed” in the cores of proteins [8].

In complementary packing, side-chains in the cores of proteins fit together without leaving any large cavities. They do so by maximizing hydrophobic

contacts while avoiding any steric clashes that could occur. As the geometric requirement of complementary packing is dependent on the detailed properties of the side-chains involved (e.g. polarity, shapes, sized), the identities of core residues would, in turn, determine the protein's high-resolution structure.

The above discussion gave a simple and straightforward view on how the folding code is being interpreted. However, we should keep in mind that things are more complicated in reality as illustrated in the following examples:

- (a) Unlike α -helix propensity, β -sheet propensity of amino acids was later found to be context dependent [10]. The use of an edge strand rather than a center strand in the same β -sheet (of IgG-binding domain from protein G) for experimentation yielded a different scale of propensities.
- (b) In a related study, Minor and Kim successfully designed a so-called "chameleon" sequence that could fold as an α -helix when in one position, but as a β -sheet when in another position of the primary sequence of the IgG-binding domain of protein G [11]. This study demonstrated that the propensity of individual amino acids to form particular secondary structures is the result of intrinsic propensity, as well as, non-local interactions. In fact, a database survey of proteins with known three-dimensional structures revealed many naturally occurring proteins with "chameleon" sequences [12].
- (c) Short, disulfide-rich peptides such as the α -conotoxin family of neurotoxic peptides all fold into the same disulfide scaffold despite hypervariability of

the primary amino acid sequences [13] (Figure 1.2A). This hypervariability did not display any conservation of binary polar and non-polar amino acid patterns that was thought to determine the global topology of a protein fold. Since all members of the α -conotoxin family possess the same cysteine framework, it had been suggested that it was the identical cysteine pattern that contributed to the common fold. However, the related χ/λ -conotoxin family of neurotoxins also had the same cysteine framework, but they fold into an alternate disulfide scaffold (Figure 1.2B).

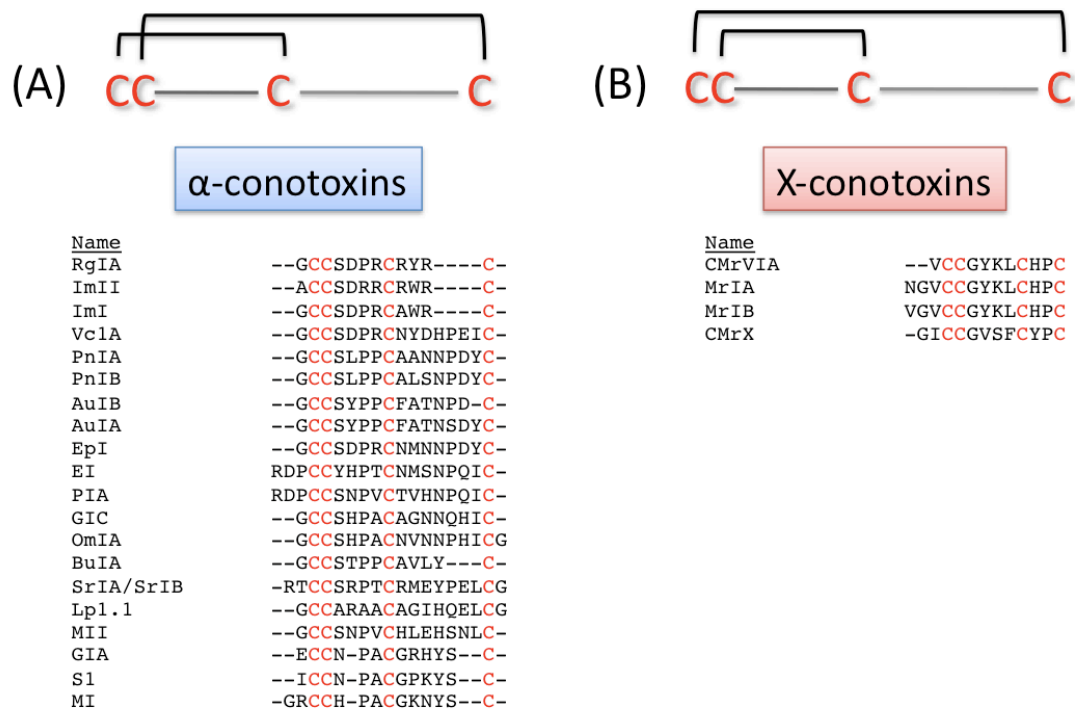


Figure 1.2 Disulfide scaffold of α - and X-conotoxins. (A) Despite hypervariability of primary amino acid sequences, without conservation of binary polar and non-polar amino acid patterns, all α -conotoxins fold into the same disulfide scaffold. (B) Despite identical cysteine patterns, α - and χ -conotoxins fold into distinct disulfide scaffolds.

All the above examples tell us that there is still a large gap in our current understanding of the mechanism behind the interpretation of the folding code.

Thus, the deciphering of the folding code still present an important field of

research despite the continual emergence of successful protein design based on variants of existing proteins and broadened alphabets of non-natural amino acids [14].

1.1.2 The folding pathway

In 1969, Cyrus Levinthal formulated the well-known Levinthal's paradox [15], a thought experiment which explained the requirement of a folding pathway. In a standard illustration of the thought experiment, the phi (Φ) and psi (Ψ) angle of each amino acid residue in a polypeptide chain is assumed to have only 3 possible conformations respectively. Accordingly, a 100-residue polypeptide chain will have a total of 198 phi/psi angles that is free to vary, resulting in a total of 3^{198} possible three-dimensional structures. If the polypeptide chain were to sample all possible structures at a rate of 10^{13} per second (or 3×10^{20} per year) before picking out the most thermodynamically stable structure to adopt, it would take approximately 10^{73} years for the 100-residue polypeptide chain to settle into its final structure. This time scale is more than astronomical if we were to take into consideration the fact that the Big Bang only occurred about 1.37×10^{10} years ago. However, the real paradox in this case lies in the empirical observation that small proteins such as the Engrailed Homeodomain protein [16] and cytochrome c [17] could fold on a microsecond to millisecond time scale.

With such a short time scale, it is reasonable to postulate the existence of specific folding "route" that leads a polypeptide chain towards its native structure, thus allowing it to by-pass structures that are irrelevant or sub-

optimal. To better understand this view, the following analogy could be used: Imagine you have to travel to London from Singapore for a business trip. If you do not have a definite path in mind, it will take you literally forever to reach London as you are just bumping around hoping to chance upon the English capital. However, if you have a definite itinerary, you could reach London in a matter of hours. What a difference in time-scale the presence of a defined pathway could make!

The necessity of a protein folding pathway has led to intense research in this area. Examples of questions that have driven this field over the years are: Does the folding of a polypeptide chain proceed in a hierarchical manner? Does protein collapse to form compact non-native structures before actual structure formation? Does a folding nuclei exist? Does folding involve only a single distinct pathway or is multiple pathways possible?

All these questions led to a multitude of possible solutions for the folding pathway puzzle. These include the “framework model”, “hydrophobic collapse model”, “nucleation-condensation model” and “energy landscape theory”. A brief description of each model is as follows:

(a) Framework model [18, 19] —

According to this model, a protein achieves its native structure in a stepwise manner, without the result of each step being re-considered at subsequent steps. Here, native secondary structures form before merging into a compact intermediate with a native-like structure. This is followed

by the formation of specific atomic interactions which will refine the tertiary structure of the protein.

(b) Hydrophobic collapse model [20, 21] —

In the hydrophobic collapse model, the polypeptide chain would first “collapse” into a more compact step before the initiation of secondary structure formation. The “collapse” is driven by the burial of hydrophobic side-chains due to the energetic stabilization conferred when they are sequestered from the surrounding water. This collapsed intermediate is also known as the “molten globule” and it is considered a “thermodynamic state” whose energy is lower than that of the denatured state but higher than that of the native state.

(c) Nucleation-condensation model [22-24] —

The nucleation-condensation model is an integration of the framework and the hydrophobic collapse model. The model describes a folding process which is analogous to crystal formation where an initial nucleation phase precedes outward crystal growth from the core. Here, a part of the polypeptide chain folds significantly earlier than other parts of the molecule, forming a nucleation site. This site, by initiating the first few correct secondary and tertiary structure interactions, then catalyzes further folding. From here, the folding reaction proceeds by having structure formation along the rest of the polypeptide chain which “condenses” or “collapses” onto the nucleation site, thus stabilizing the nucleus of the protein.

(d) Energy landscape theory [25, 26] —

Unlike other models of the protein folding pathway, the energy landscape theory assumes that folding occurs through organizing an ensemble of structures rather than through uniquely defined structural intermediates. Specifically, is a statistical description of a protein's potential surface where a rugged funnel-like energy landscape biased the folding polypeptide towards its native structure. The mouth of the funnel represents the large entropy of the denatured state — i.e. A large ensemble of denatured structures with high energy. As native/favorable contacts are formed, the stabilization energy will decrease with a concomitant drop in configurational entropy. This then pushes the folding polypeptide towards the single lowest energy structure which will become its native conformation.

1.2 Disulfide Bonds as Probes of Protein Folding

1.2.1 Trapped disulfide-containing intermediates for the study of protein folding pathway

Techniques such as fluorescence spectroscopy, pressure-jump relaxation, temperature-jump relaxation, hydrogen exchange pulse labeling and stopped-flow circular dichroism had been used to study protein folding dynamics. Although these techniques allow the observation of protein folding events in the microsecond to millisecond timescale, they do not allow folding intermediates to be isolated for detailed characterization. This is due to the fact that folding intermediates are thermodynamically unstable and thus do not accumulate significantly at equilibrium for them to be characterized. However, if these folding intermediates could somehow be trapped or “frozen” in time, it would offer a solution to the problem. To this end, disulfide bond-containing proteins had been suggested to be good candidates for detailed characterization of folding intermediates as they could be chemically trapped in the course of folding [27]. This is due to the unique chemistry of cysteine residues which are involved in disulfide bond formation.

The folding of proteins containing disulfide bonds consist of two interdependent processes: (1) conformational folding and (2) disulfide bond regeneration [28]. During the course of conformational folding, two thiol groups (of cysteine residues) which are in close proximity to each other might form a disulfide bond via rearrangement (i.e. disulfide shuffling) or oxidation (e.g. air oxidation). Any free thiols present in the protein at this time could be chemically modified by iodoacetamide to prevent further disulfide bond

generation and thus conformational folding [29]. Another way to pause oxidative folding is by acid trapping — To acidify the folding solution to disfavor the deprotonation of thiols to thiolates, the active species involved in disulfide bond formation [30].

Trapped disulfide-containing folding intermediates had been used to elucidate the folding pathways of bovine pancreatic trypsin inhibitor (BPTI) [31, 32], hirudin [33], epidermal growth factor (EGF) [34], leech carboxypeptidase inhibitor (LCI) [35], α -lactalbumin [36] and RNase A [37]. The folding pathways of these disulfide-rich proteins had provided supporting evidence for the existence of various pathways suggested by fast-kinetic studies — For example: In BPTI, a limited number of native-like intermediates funnel the protein towards its native structure, thus making this kind of folding in line with the “framework model” where local interactions is important in guiding the protein through the hierarchic condensation of native-like elements. On the other hand, hirudin-like proteins fold through an initial stage of disulfide bond formation followed by the rearrangement of isomers to form the native protein, thus making this kind of folding in line with the “hydrophobic collapse model” where an initial stage of collapse is followed by a slower annealing phase in which specific interactions are used to refine the structure [38].

1.2.2 Disulfide-connectivity based structural isoforms for the study of protein folding code

In Section 1.1.2, we concluded that the absence of a folding pathway would require a 100-residue polypeptide chain to sample through 3^{198} possible

conformations (if each ϕ - and ψ -angle is given 3 degrees of freedom) to find its native conformation. However, this is on the assumption that no cysteine residues, which could potentially participate in disulfide bonds, is present. Due to the structural constraint conferred by disulfide bonds, the inclusion of 6 cysteine residues in the amino acid sequence of the polypeptide chain (i.e. 6 cysteines, 94 non-cysteines) will make 3 disulfide bonds (if fully oxidized), resulting in the reduction of possible conformations to only 15. A conformational space of 3^{198} *versus* 15 makes a 93 order of magnitude difference! Even if it contains 17 disulfide bonds, the conformational possibilities of 6.33×10^{18} will still be 76 order of magnitude lower than without any disulfide at all. Thus, the formation of disulfide bonds, whether native or non-native, during the process of protein folding is another innovative way to minimize the conformational search of a polypeptide chain.

The presence of non-native disulfide bonds in folding intermediates seemed peculiar. However, it is not — Instead of directing folding in the wrong direction, structural constraints imposed on the folding intermediates by non-native disulfide bonds had been suggested to enhance the folding process by creating a compact fold, thus bringing other cysteine residues and different parts of the polypeptide chain into close proximity to facilitate the re-shuffling of disulfide bonds and the concomitant formation of the native structure, respectively [39].

However, to guarantee the success of this useful strategy, nature must place sufficient information in the protein folding code to ensure that the disulfide-connectivity is correct at the end of the folding process. If not, structural

constraint exerted by incorrect disulfide-connectivity might “lock” the protein into an incorrect conformation, thus negating any positive effects the formation of disulfide bonds have on the folding process.

The presence of such information in the protein folding code is exemplified in *in vitro* oxidative folding experiments using atmospheric oxygen as the oxidizing agent. Unlike the use of redox buffer system, such as cysteine/cystine or reduced/oxidized glutathione, where disulfide bonds could be continually reduced then re-formed, cysteine oxidation by atmospheric oxygen goes through free radical intermediates [40] which is irreversible once all thiol groups available had been engaged in disulfide bonds (i.e. no disulfide shuffling). In spite of this, fully reduced proteins or peptides such as ribonuclease [1-3] and α -Conotoxin Iml [41] respectively, had been shown to recover their native disulfide-connectivity in reasonable yield upon re-oxidation by atmospheric oxygen. These results demonstrated that information for correct disulfide-connectivity is encoded in the primary amino acid sequence of the protein itself (as anticipated by Anfinsen’s thermodynamic hypothesis) and the “instructions” given by them is needed to form native disulfide bonds in the presence of other highly competitive oxidative processes.

In view of the above discussion, one can see that the dictation of correct disulfide-connectivity is an integral part of the protein folding code. Thus, it is important for us to understand how this information is being embedded in the amino acid sequence of a protein. A good model to use for the understanding of native disulfide-connectivity determination is that of short peptides containing four cysteine residues. By careful manipulation of sequence

information and/or oxidative folding conditions, one could pick out structural determinants that influence the disulfide-connectivity choices. The limited subset of structural isoforms in these simple models (i.e. 3 isoforms for 2 disulfide bonds) allows us to see the influence of minor manipulations on folding tendency quantitatively.

1.3 The Canonical Fold of the EGF-like domain

In light of the gaps still present in our knowledge of the protein folding code (Section 1.1.1), this study was undertaken to provide more insights into this aspect of the protein folding problem. To this end, the canonical fold of the evolutionarily conserved epidermal growth factor (EGF)-like domain was chosen as the subject of our study.

1.3.1 Description of the canonical EGF-like domain fold

The EGF-like domain is a sequence of about 30 to 40 amino acid residues, with the epidermal growth factor itself being the prototype sequence [42] (Figure 1.3A). A notable feature of all EGF-like domains is the evolutionary conservation of six cysteine residues in defined positions along the amino acid sequence as well as a glycine and aromatic residue in the third inter-cysteine region (Figure 1.3B).

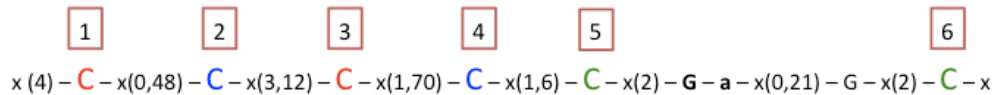
With regards to the secondary and tertiary structure of the canonical EGF-like domain, it folds into a three-looped structure made up of a central two-stranded β -sheet followed by a loop to a short C-terminal two-stranded sheet (Figure 1.4). This structure is stabilized by three disulfide bonds formed between the first and third, second and fourth, fifth and sixth cysteine residue (C_1 - C_3 , C_2 - C_4 , C_5 - C_6)¹ of the domain.

¹ Annotation of disulfide-connectivity: "C" denotes a cysteine residue. Number in subscript represents the relative position of the cysteine residue along the amino acid sequence from the N-terminal to C-terminal — e.g. "1" means first cysteine residue, "2" means second cysteine residue, "3" means third cysteine residue, etcetera. "-" (Dash) denotes the connectivity between the two indicated cysteine residues.

(A) Epidermal growth factor (*Mus musculus*)



(B) EGF-like domain (Consensus sequence)



✧ 'C': conserved Cys involved in a disulfide bond
 ✧ 'G': often conserved glycine
 ✧ 'a': often conserved aromatic amino acid
 ✧ 'x': any residue

Figure 1.3 The consensus sequence of the EGF-like domain. (A) The epidermal growth factor serves as the prototype sequence of the EGF-like domain. Its disulfide-connectivity is indicated by square brackets connecting the respective cysteine residues. (B) The consensus sequence of the EGF-like domain.

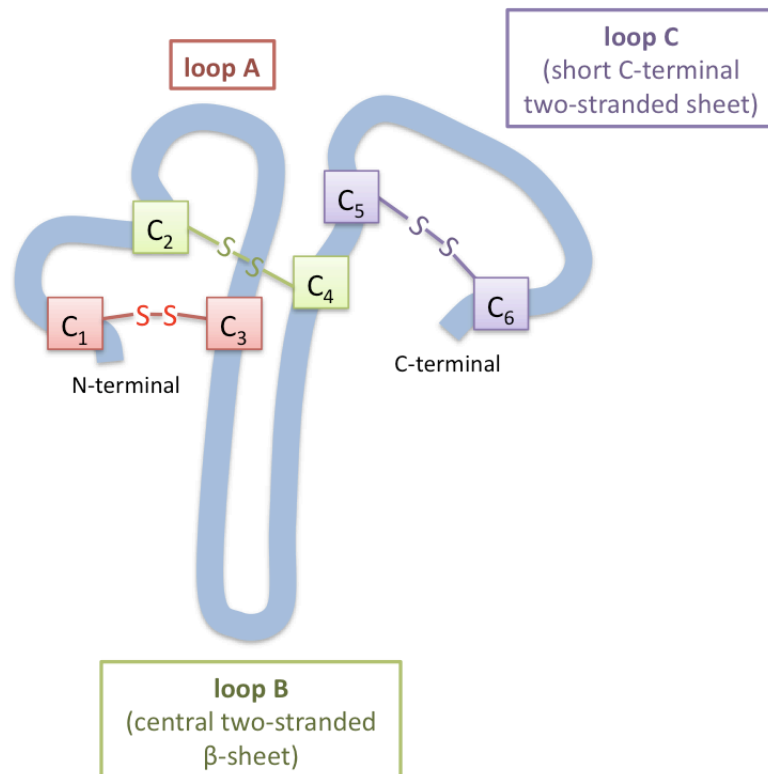


Figure 1.4 The canonical fold of the EGF-like domain. It consists of 3 loops, labeled as loop A, loop B and loop C, respectively. Cysteine residues are numbered according to their relative position along the amino acid sequence from the N-terminal to C-terminal. The three disulfide bridges are also indicated.

1.3.2 Significance of studying the protein folding code of EGF- like domain

EGF-like domains are found in the extracellular domain of membrane-bound proteins or in secreted proteins. They have been the subject of many biological investigations because it is an evolutionarily conserved protein domain with diverse functions — For example, EGF-like domains from various proteins had been shown to be capable of: (a) Mediating receptor-binding for host-cell recognition in parasitic infection [43] ; (b) Conferring functional differences (activator or inhibitor) to various ligands involved in receptor signaling during embryogenesis [44] ; (c) Binding to calcium ions² which serves to orient neighboring modules relative to each other in a manner that is required for biological activity (e.g. factor IX — Gla-EGF fragment) [45]. Of course, the above-mentioned functions are only a tiny fraction of the vast functional capabilities of the EGF-like domains, but as this aspect of EGF-like domain biology is beyond the scope of this thesis, a detailed description shall not be attempted.

However, of considerable interest in our discussion here is how do protein domains like the EGF-like domains achieve such an array of functional diversity? One reasonable explanation would be that of domain duplication during evolution, followed by accumulated amino acid changes in the duplicated domain to generate functional diversity. Indeed, functional divergence of the EGF-like domain had lead to hypervariability of amino acid

² Constitute a distinct subset of EGF-like domain. The consensus sequence for calcium binding is D/N-x-D/N-E/Q-y-D/N-y-Y/F, where x indicates a variable amino acid and y indicates a sequence of variable amino acids.

sequence in its inter-cysteine region.

Despite sequence hypervariability, most EGF-like domains (based on those with structures solved) fold into the canonical three-looped structure that is defined by a disulfide-connectivity of C₁-C₃, C₂-C₄, C₅-C₆ (Figure 1.4). For this to be possible, the perseverance of folding information in the amino acid sequence is necessary while functional evolution is taking place. However, the exact nature of this folding information is currently unknown — Among the 30 to 40 amino acid residues of the EGF-like domains, which are the “functional” residues and which are the “structural” residues? The “structural” residues constitute the protein folding code and they dictate the native three-dimensional structure of the domain. This view slightly deviates from the traditional concept of the protein folding code in which the amino acid sequence in its totality determine the native structure of the protein. Here, only structural determinants are needed and they are interspersed in the amino acid sequence together with residues needed for the functional capability of the protein. This way of organizing “structure-function” information in the amino acid sequence allows for functional diversity to develop on a single protein scaffold.

Here, the study of the folding code of the canonical fold of the EGF-like domain serve as a good starting point to provide more insights into the nature of structural determinants — What are they, where are they located in the amino acid sequence and the mechanism by which they act. The focus of this study would be at the level of disulfide-connectivity as it is this aspect of the EGF-like domain structure that is most conserved despite slight variation in

the structure of the inter-cysteine loops³. In the case of EGF-like domain, the conservation of disulfide-connectivity had also led to the conservation of the overall fold, thus the study of the disulfide-connectivity is in itself a useful probe to understand the folding code of this domain.

³ Slight variation in the structure of the inter-cysteine loops is inevitable due to the hypervariability of amino acid sequence in these regions of the EGF-like domains. However, it is important to note that only exact structural details are affected, while the overall three-looped structure is maintained throughout all EGF-like domains.

1.4 Thrombomodulin and Its Role in the Anti-coagulation Pathway

To determine the structural determinants that dictate the canonical fold (i.e. disulfide-connectivity) of the EGF-like domain, the fourth and fifth EGF-like domain of thrombomodulin (TM) were chosen as models for the study.

Here, the role of TM in the anti-coagulation pathway will be discussed to aid in the understanding of the structural significance regarding its smallest co-factor active fragment — TM EGF-like domain 4 and 5.

1.4.1 TM as a regulator of the coagulation cascade

1.4.1.1 Thrombin: The coagulant

During secondary hemostasis, proteins in the blood plasma, called coagulation factors, engage in a complex pathway to form a fibrin meshwork. The purpose of this fibrin meshwork is to strengthen the platelet plug, which is formed during primary hemostasis, at the site of blood vessel injury.

Thrombin, also known as coagulation factor II (FII), is a serine protease which acts as the direct effector in the formation of the fibrin meshwork. It does so in two steps: First, by converting fibrinogen into fibrin [46, 47] with the concomitant self-polymerization of fibrin monomers [48-50]. Second, by activating factor XIII (FXIII) into FXIIIa [51] which is responsible for the covalent cross-linking of the established fibrin-polymer [52].

In addition to its direct effects on fibrin meshwork formation, thrombin has the ability to amplify its own generation via the activation of other coagulation factors. Thrombin does so by activating:

- (a) Factor XI (FXI) into FXIa [53] which, in turn, activates factor IX (FIX) into FIXa [54].
- (b) Factor VIII (FVIII) into FVIIIa [55] which, in turn, acts as a cofactor for FIXa [56]. Together, they form the intrinsic tenase complex which activates factor X (FX) into FXa [57].
- (c) Factor V (FV) into FVa [58] which, in turn, acts as a cofactor for FXa [59]. Together, they form the prothrombin complex which activates prothrombin into thrombin, resulting in a positive feedback loop.

1.4.1.2 Thrombin-TM complex: The anti-coagulant

Although thrombin is an effective coagulant, this function could be reversed by the binding of TM, a transmembrane glycoprotein expressed on the luminal surface of vascular endothelial cells, in a 1:1 stoichiometric complex. By acting as a cofactor, TM serves as a molecular switch that turns thrombin into an anti-coagulant [60, 61].

The thrombin-TM complex exerts its anti-coagulant activities in two main ways: (a) Passively, by preventing the binding of thrombin's pro-coagulant substrates (fibrinogen, FV and FVIII [62-65]) ; (b) Actively, by the activation of protein C [66] which is a serine protease like thrombin. When protein C is activated by the thrombin-TM complex, it goes on to inactivate FVIIIa (with protein S and intact FV as cofactors) and FVa (with protein S as cofactor) [67-69], thus shutting down the thrombin-mediated positive feedback loop on its own activation, and the reduction in formation of fibrin from fibrinogen.

From the above discussion, we can see that TM plays a central role in the

homeostasis of the blood coagulation system by making thrombin a pivoting factor between pro-coagulation and anti-coagulation. Indeed, the importance of TM in the blood coagulation system could be exemplified by arterial and venous thrombotic diseases caused by mutations in the TM gene which resulted in the reduced expression of the TM protein [70]. In addition, effects of TM gene mutation could also be seen at the level of embryonic development — Isermann et al. showed that disruption of the mouse TM gene led to embryonic lethality due to activation of the blood coagulation at the fetomaternal interface which resulted in the death of trophoblast cells [71].

1.4.2 Structure-function relationship of TM: Role of the fourth to sixth EGF-like domains

TM is a multi-modular protein consisting of a lectin-like domain at the amino terminal, followed by a hydrophobic segment, six tandem EGF-like domains, an *O*-glycosylated serine/threonine-rich domain, a trans-membrane segment and a short cytoplasmic tail (Figure 1.5).

The smallest cofactor active fragment of TM had been identified as the fourth and fifth EGF-like domains. Together they constitute 10% of the specific activity of TM which is greatly enhanced when the sixth EGF-like domain is included [72]. Studies involving the individual EGF-like domain that constitute the cofactor active fragment had given us useful insights into the function of each specific domain:

The fourth EGF-like domain (TM EGF D4) alone did not display any cofactor

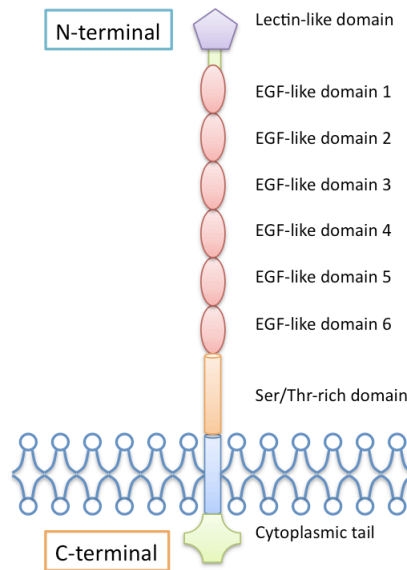


Figure 1.5 The domain organization of thrombomodulin.

activity when assayed as a replacement for full-length TM in protein C activation assay. It also did not display any ability to bind to thrombin when assayed as a competitive inhibitor for protein C activation when full-length TM is included in the reaction [73]. On the other hand, a TM fragment consisting of the fifth and sixth EGF-like domain (TM EGF D5-D6) was shown to bind to thrombin with high affinity by being a competitive inhibitor of thrombin-TM in the activation of protein C. However, like TM EGF D4, this fragment alone did not any cofactor activity [74]. These results support the view that although TM EGF D5 and TM EGF D6 could bind to thrombin, it needs TM EGF D4 for cofactor activity. Meanwhile, TM EGF D4 could not exert its function as it could not associate with thrombin without the help of TM EGF D5-D6. Further support for the central role of TM EGF D4-D6 in TM's function comes in terms of structural evidence provided by Fuentes-Prior et al. [75]:

In a 2.3 Å crystal structure of human α -thrombin bound to the TM EGF D4-D6

fragment (Figure 1.6), it was demonstrated that TM EGF D5 and part of TM EGF D6 bind to a cluster of lysine and arginine residues in the anion-binding exosite-I of thrombin. Since thrombin's procoagulant substrates like fibrinogen, FV and FVIII also bind to thrombin via exosite-I [62-65], the competitive binding of TM EGF D5-D6 segment to the same site provides the basis of blockade of procoagulant substrates in the thrombin-TM complex.

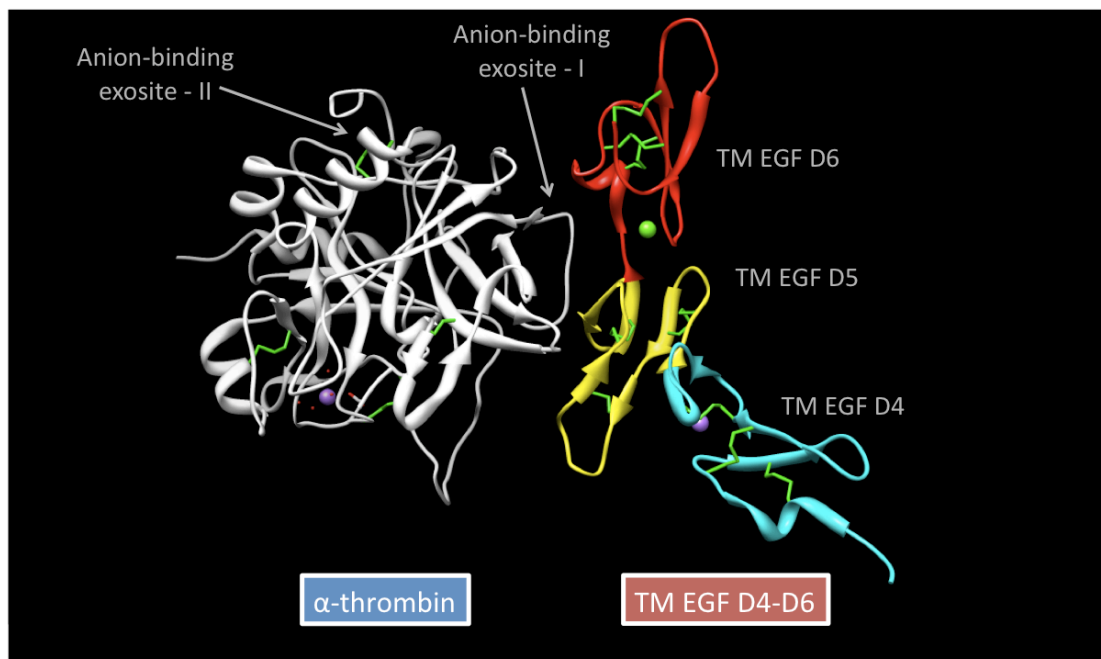


Figure 1.6 Ribbon model of the complex between α -thrombin and TM EGF D4-D6 [PDB: 1DX5]. α -Thrombin is shown in white. TM EGF D4, TM EGF D5 and TM EGF D6 are shown in cyan, yellow and red, respectively. Disulphide linkages are shown in green.

On the hand, the TM EGF D4 segment was shown to be anchored almost perpendicular to the linear TM EGF D5-D6 tandem. It protrudes away from thrombin, and thus does not interact directly with it. It was suggested that thrombin binding to the TM EGF D5-D6 segment creates an additional substrate-binding interface on TM EGF D4-D5. The “free” TM EGF D4 segment is then needed to interact with anti-coagulant substrates of thrombin (i.e. protein C) such that it positions the scissile peptide bond of the substrate

with the catalytic machinery of thrombin in an optimal stereochemical conformation for cleavage. This interaction provides the structural basis for the alteration of thrombin's substrate specificity upon TM's binding.

1.5 Thrombomodulin EGF-like Domain 4 and 5: Models in the Study of the EGF-like Domain Folding Code

1.5.1 TM EGF D4 *versus* TM EGF D5: Canonical *versus* non-canonical EGF-like domain fold

As TM EGF D4-D5 is the smallest active cofactor of TM, there had been keen interest in solving the three-dimensional structure of these two domains as part of a larger effort to understand the structure-function relationship of TM. The results of these research had led to the discovery of a non-canonical EGF-like domain fold involving TM EGF D5 [76-78]. This non-canonical structure of EGF-like domain is defined by a different disulfide-connectivity, and thus, is of considerable interest in this current study regarding the folding code of the canonical EGF-like domain. Below will be a brief description of the structure of TM EGF D4 and D5 with the intent to highlight the key differences between these two domains with regards to their three-dimensional structure.

1.5.1.1 Solution structure of TM EGF D4

The solution structure of human TM EGF D4 has been determined by 2D ¹H NMR [73, 79]. Here, the overall structure resembles that of the canonical EGF-like domain. Residues that are important for cofactor activity (Glu357, Tyr358, Gln359, Glu374 and Phe376), as determined by alanine scanning experiments, are found to form a “patch” that is exposed to solvent in the structure of TM EGF D4 [73]. More importantly (due to the purpose of this study), TM EGF D4 possesses the canonical EGF-like domain disulfide-connectivity of C₁-C₃, C₂-C₄, C₅-C₆ (Figure 1.7).

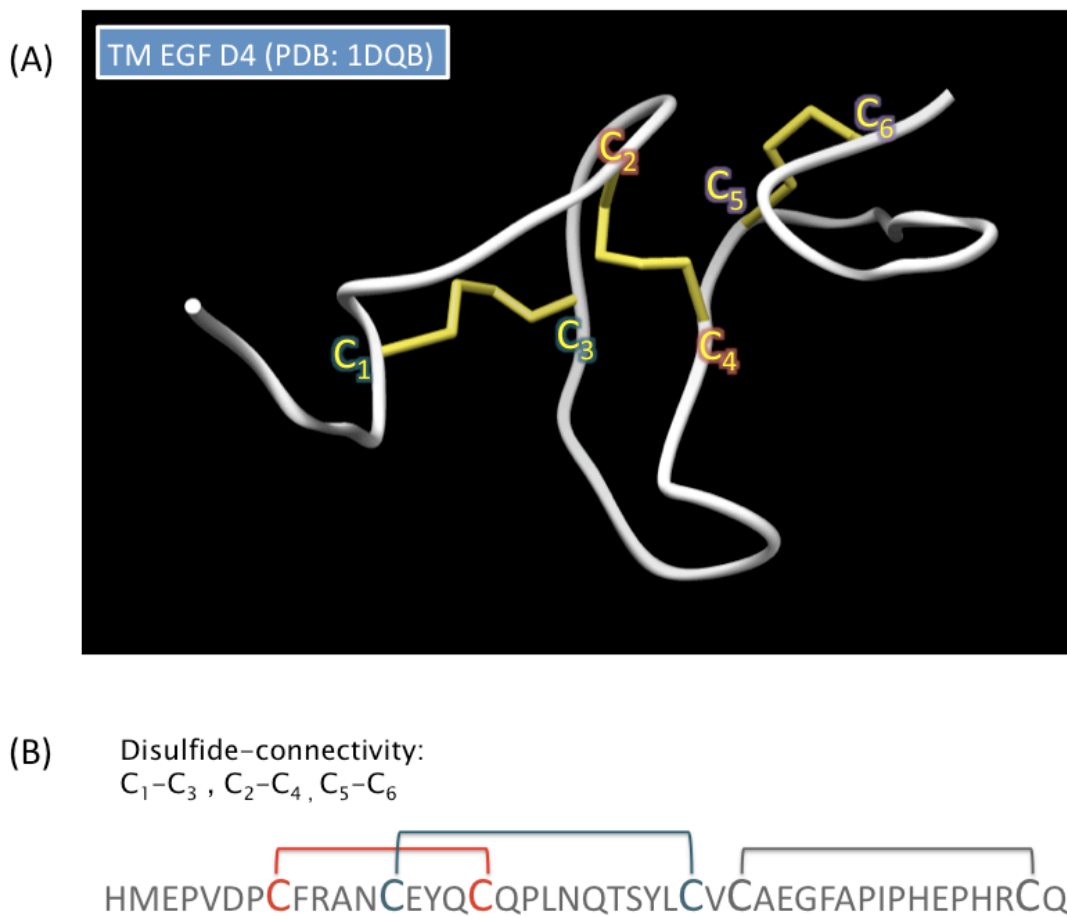


Figure 1.7 Solution structure of TM EGF D4 and its disulfide-connectivity. (A) Backbone structure of TM EGF D4 (white) [PDB: 1DQB]. Disulfide linkages are indicated in yellow. Locations of cysteine residues are labeled as C_1 , C_2 , C_3 , C_4 , C_5 and C_6 . (B) The amino acid sequence of TM EGF D4 with disulfide-connectivity indicated. The disulfide-connectivity of TM EGF D4 is C_1-C_3 , C_2-C_4 , C_5-C_6 . This is the disulfide-connectivity of the canonical EGF-like domain.

1.5.1.2 Solution structure of TM EGF D5

Like TM EGF D4, the structure of human TM EGF D5 has also been determined by 2D ^1H NMR [76, 79]. The structure of this domain appears to have diverged from the canonical EGF-like structure — The central two-stranded β -sheet in the canonical EGF-like domain is absent in TM EGF D5. Furthermore, the N- and C-termini is closer together in TM EGF D5 than in other EGF-like domains. In addition to structural divergence from the

canonical EGF-like fold, it is important to note that TM EGF D5 also possess a novel disulfide-connectivity of C₁-C₂, C₃-C₄, C₅-C₆ (Figure 1.8).

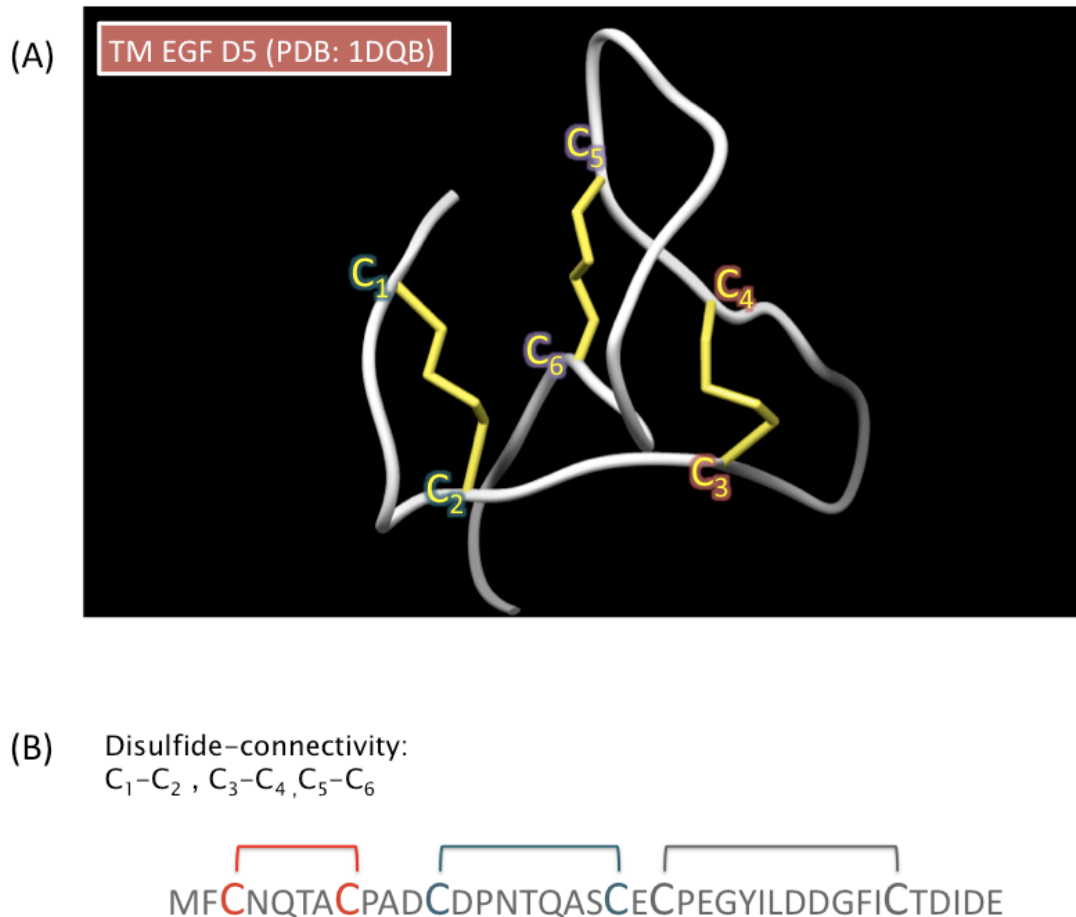


Figure 1.8 Solution structure of TM EGF D5 and its disulfide-connectivity. (A) Backbone structure of TM EGF D5 (white) [PDB: 1DQB]. Disulfide linkages are indicated in yellow. Locations of cysteine residues are labeled as C₁, C₂, C₃, C₄, C₅ and C₆. (B) The amino acid sequence of TM EGF D5 with disulfide-connectivity indicated. The disulfide-connectivity of TM EGF D4 is C₁-C₂, C₃-C₄, C₅-C₆. This disulfide-connectivity is non-canonical.

The functional significance of this unique non-canonical structure (and disulfide-connectivity) can be seen from a series of experiments performed by Meininger, Hunter and Komives [78]. In these experiments, various structural isoforms of TM EGF D5, based on differential disulfide-connectivity, were tested for thrombin-binding affinities through two kinds of thrombin inhibition assays — (a) Amount of peptide needed to double fibrinogen clotting time

(thrombin inhibition) and (2) inhibition of protein C activation (competition with native full-length TM for thrombin binding). Key results from these two assays are highlighted below (Table 1.1):

Table 1.1 *Effect of various TM EGF D5 structural isoforms on thrombin activity*

TM EGF D5 Structural Isoform	Amount of peptide to double clotting time (μM)	K_i for protein C activation (μM)
$C_1\text{-}C_2, C_3\text{-}C_4, C_5\text{-}C_6$	210 ± 50	370 ± 50
$C_1\text{-}C_3, C_2\text{-}C_5, C_4\text{-}C_6$	340 ± 50	830 ± 50
[xt] $C_1\text{-}C_2, C_3\text{-}C_4, C_5\text{-}C_6$	0.2 ± 0.02	1.9 ± 0.2
[xt] $C_1\text{-}C_3, C_2\text{-}C_4, C_5\text{-}C_6$	9 ± 1	13 ± 1

Note. Structural isoforms are defined by disulfide-connectivity. [xt] denotes an extended form of the TM EGF D5 isoform which included four additional amino acids connecting the fifth and sixth EGF-like domains of TM. Adapted from "Thrombin-binding affinities of different disulfide-bonded isomers of the fifth EGF-like domain of thrombomodulin," by M.J. Hunter and E.A. Komives, 1995, *Protein Science*, 4, p. 2134.

From these results, it is apparent that the $C_1\text{-}C_2, C_3\text{-}C_4, C_5\text{-}C_6$ isoform of TM EGF D5 is a better inhibitor of thrombin activity and thrombin-TM interaction than the $C_1\text{-}C_3, C_2\text{-}C_5, C_4\text{-}C_6$ isoform. This is because a lower amount of $C_1\text{-}C_2, C_3\text{-}C_4, C_5\text{-}C_6$ isoform is needed to achieve comparable inhibition level in both assays. Moreover, for the extended isoforms of TM EGF D5, which included four additional amino acids from the linker region between TM EGF D5 and TM EGF D6 (for the purpose of better binding), the [xt] $C_1\text{-}C_2, C_3\text{-}C_4, C_5\text{-}C_6$ isoform could inhibit both thrombin activity and thrombin-TM interaction better than the [xt] $C_1\text{-}C_3, C_2\text{-}C_4, C_5\text{-}C_6$ (canonical EGF-like domain) isoform. Therefore, the non-canonical structure/disulfide-connectivity of TM EGF D5 has a high structure-function significance. This indicates that this highly

divergent EGF-like domain had been evolutionarily selected for, and is not simply a neutral mutation which has been “accidentally” preserved.

1.5.2 TM EGF D4 and TM EGF D5 as models to identify the structural determinants of the canonical EGF-like domain fold

Due to the different nature of TM EGF D4 and D5 with respect to their structure (i.e. TM EGF D4 is canonical EGF-like, while TM EGF D5 is non-canonical), they serve as contrasting models in the study of the EGF-like domain folding code. Moreover, since the structural divergence of TM EGF D5 had been evolutionarily selected for, it also serves as an interesting model to show how structural divergence can be achieved within a single domain.

In the context of our study, the different native disulfide-connectivity of TM EGF D4 and TM EGF D5, with corresponding difference in structures, serve as a useful tool for the identification of structural determinants in the canonical EGF-like domain (C_1-C_3 , C_2-C_4 , C_5-C_6) fold. The criteria for being the structural determinants based on these two contrasting models are:

- (a) The amino acid qualifying as the structural determinant of TM EGF D4 should not be present at its equivalent position in TM EGF D5, and *vice versa*.
- (b) When the structural determinant of TM EGF D4 is replaced with another residue of different physical-chemical property, it should change its folding tendency to that of the non-canonical fold.

(c) When TM EGF D4's structural determinant is placed into its equivalent position in TM EGF D5, it should increase TM EGF D5's folding tendency towards that of the canonical EGF-like domain fold.

These criteria are based on the hypothesis that the switch from the canonical C₁-C₃, C₂-C₄ conformer to the non-canonical C₁-C₂, C₃-C₄ conformer is the result of a change in the physical-chemical properties of the canonical fold's structural determinants. The change in physical-chemical properties of the structural determinants will then be manifested as a change in the dominant force of folding, thus resulting in a different final structure and disulfide-connectivity. Therefore, studies that quantify the relative contribution of various inter-molecular forces to the folding tendencies of both domains would provide the clue to the nature and identity of the structural determinants.

1.6 Objectives and Scope of the Thesis

The main objective of this thesis is to identify structural determinants that are responsible for dictating the alternate disulfide-connectivity of TM EGF D4 and TM EGF D5. The scope of this thesis covers the following areas:

(a) General localization of the structural determinants in TM EGF D4 and D5.

More specifically, it is to find out whether their respective structural determinants are located locally within the segment encompassing C₁ to C₄ (where disulfide-connectivity difference of TM EGF D4 and D5 lies) or if the C-terminal segment of the domain (encompassing C₅ to C₆) has a role in influencing the different disulfide-connectivity preference of the front segment.

(b) Determination of the dominant force that dictates the folding tendency/ disulfide-connectivity preference of each domain (i.e. hydrophobic or electrostatic).

(c) Identification of key residues as structural determinants in TM EGF D4 and D5. This would be aided by knowledge regarding the general localization of the structural determinants and the nature of the dominant force that drive their respective folding tendencies.



Chapter 2: Materials and Methods



2.1 Peptide Synthesis and Purification

2.1.1 Peptide synthesis

Peptides were synthesized using manual 9-fluorenylmethoxycarbonyl (Fmoc)-solid phase peptide synthesis. All amino acids used were Fmoc-L-(amino acid)-OH derivatives, with some residues containing side-chain protection groups. The side-chain protected amino acids used were: Arg(Pbf), Asn(Trt), Asp(OtBu), Cys(Trt), Cys(Acm), Gln(Trt), Glu(OtBu), His(Trt), Ser(tBu), Thr(tBu), and Tyr(tBu). For synthesis of truncated TM EGF D4 and D5 structural isoforms (t-TM EGF D4 and t-TM EGF D5), Cys(Trt) and Cys(Acm) were incorporated at specific locations along the amino acid sequence — (a) C₁: Cys(Trt), C₂: Cys(Acm), C₃: Cys(Trt), C₄: Cys(Acm) for C₁-C₃, C₂-C₄ isoform ; (b) C₁: Cys(Trt), C₂: Cys(Trt), C₃: Cys(Acm), C₄: Cys(Acm) for C₁-C₂, C₃-C₄ isoform ; (c) C₁: Cys(Acm), C₂: Cys(Trt), C₃: Cys(Trt), C₄: Cys(Acm) for C₁-C₄, C₂-C₃ isoform (Figure 2.1). For peptides used for *in vitro* oxidative folding experiments, only Cys(Trt) was used.

The peptides were assembled on the Novasyn® TGR resin (Novabiochem, Darmstadt, Hesse, Germany), which was designed for the synthesis of peptide amides. The coupling step was performed in *N,N*-dimethylformamide (DMF): *N*-Methyl-2-pyrrolidone (NMP) (2:1) with 5 times excess of amino acid derivatives activated *in situ* by 4.9 times excess of *O*-(7-Azabenzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate (HATU) and 10 times excess of *N,N*-diisopropyl-ethylamine (DIPEA). Removal of Fmoc-moiety (deblocking) was achieved using a solution of 20% (v/v) piperidine in DMF. The

success of coupling and de-blocking was verified for each residue using the Kaiser test [80] (all amino acid residues except Pro) and Chloranil test [81] (Pro residue).

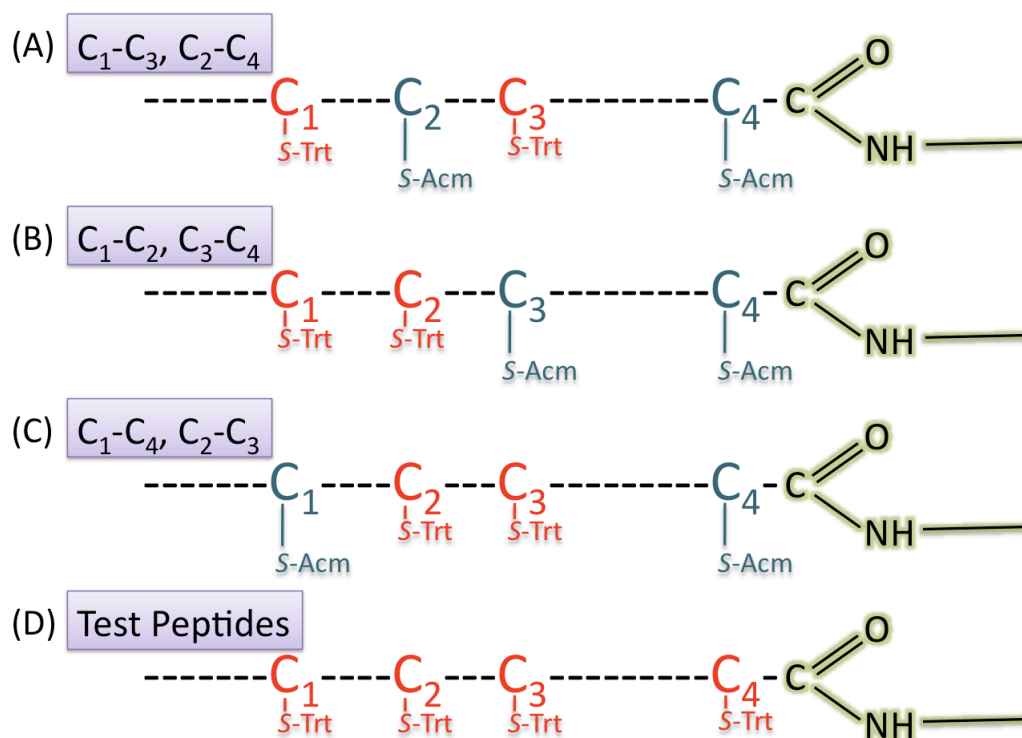


Figure 2.1 Synthesis of t-TM EGF D4 and t-TM EGF D5 structural isoforms and their respective "test peptides" using Cys(Acm) and Cys(Trt). For the synthesis of structural isoforms, Cys(Trt) and Cys(Acm) were incorporated at specific positions along the polypeptide chain as illustrated in: (A) C₁-C₃, C₂-C₄, (B) C₁-C₂, C₃-C₄, and (C) C₁-C₄, C₂-C₃. (D) For the synthesis of peptides used for *in vitro* oxidative folding experiments, only Cys(Trt) was used.

2.1.2 Peptide cleavage, deprotection and isolation

After synthesis was complete, the resin was rinsed extensively with 3 cycles of successive methanol (MeOH), DMF and dichloromethane (DCM) washes, followed by a final MeOH rinsing step before drying overnight under vacuum. Peptides without Cys(Acm) derivatives were deprotected and cleaved from the resin using a cocktail of trifluoroacetic acid (TFA)/1,2-ethanedithiol (EDT)/thioanisole/water (90:4:4:2 % v/v) for 2 hrs with gentle stirring. Peptides with Cys(Acm) derivatives were deprotected and cleaved from the resin with a

cocktail of TFA/EDT/triisopropylsilane (TIS)/water (94:2.5:1:2.5 % v/v) instead. After removal of the resin by filtration through fritted glass funnels, the peptides were precipitated by dropping the filtrate drop-wise into ice-cold diethyl-ether. The precipitate was collected as a pellet after centrifugation and allowed to dry overnight.

2.1.3 Peptide purification

Dried peptides were dissolved using 0.1% (v/v) TFA in 10% (v/v) acetonitrile (ACN) and purified using reversed-phase HPLC with a Jupiter Proteo, 4 μ , 90 Å (15 × 250 mm) column (Phenomenex, Torrance, California, USA) on an ÄKTA™ purifier system (GE Healthcare, Uppsala, Sweden). A segmented gradient elution method involving TFA as the counter-ion (constant concentration of 0.1% v/v), and ACN as the organic modifier (maximum 80% v/v) was used. The purified peptides were verified using electrospray ionization-mass spectrometry (Section 2.1.4) before lyophilization.

2.1.4 Electrospray ionization-mass spectrometry (ESI-MS)

Peptide mass determination using ESI-MS was performed on an API-300 LC/MS/MS system (Perkin-Elmer Sciex, Selton, Connecticut, USA). The samples were introduced via direct injection. The LC-10AD liquid chromatography system (Shimadzu, Kyoto, Japan) was used as the solvent delivery system with 0.1% (v/v) formic acid in 50% ACN as the solvent. Ionspray, orifice and ring voltages were set at 4600 V, 50 V and 350 V, respectively. Nitrogen was used as the nebulizer and curtain gas.

2.2 Regioselective Synthesis of Structural Isoforms

By placing *S*-trityl (Trt) or *S*-acetamidomethyl (Acm)-protected cysteine residues at specific positions along the peptide chain (Section 2.1.1), orthogonal protection of cysteine residues' side-chains were used to generate structural isoforms based on differential disulfide-connectivity. Cysteine residues involved in the formation of the first disulfide bridge were protected with the acid labile Trt-group, which were removed upon TFA treatment in the peptide synthesis cleavage step (Section 2.1.2). After formation of the first disulfide bridge, the remaining two Acm-protected cysteine residues would be treated with iodine to achieve simultaneous removal of Acm-group and oxidation to form the second disulfide bridge.

2.2.1 Formation of the first disulfide bridge

2.2.1.1 DMSO-mediated oxidation

Fully reduced, purified Cys(Acm)-containing peptides with two free cysteine residues were dissolved at a concentration of 0.3 mM in a 0.1 M Tris-HCl, pH 7.5 buffer containing 10% ACN and 20% DMSO. DMSO-mediated oxidation was allowed to take place under vigorous stirring and the progress of the reaction was monitored using the Ellman's test (Section 2.2.1.2). When reaction was completed, as indicated by a negative Ellman's test, the pH of the solution was adjusted to pH 2 using concentrated HCl. The peptide, now containing one disulfide-bridge, was directly injected into the Jupiter Proteo, 4 μ , 90 Å (15 × 250 mm) column for purification using the segmented gradient

elution method described in Section 2.1.3. The purified peptides were verified using ESI-MS (i.e. mass reduction of 2 Da) before lyophilization.

2.2.1.2 Ellman's Test

A reaction buffer of 0.1 M sodium phosphate, pH 8.0, containing 1 mM EDTA was prepared. This was followed by an Ellman's reagent solution which was made by dissolving 5,5'-dithio-bis-(2-nitrobenzoic acid) (DTNB) in the reaction buffer at a concentration of 0.4% (w/v). The proportion of Ellman's reagent, peptide sample and reaction buffer used in the test is 1:5:50, respectively. The reaction mixture was incubated at room temperature for 15 mins before the absorbance of the sample was measured at 412 nm using a NanoVue spectrophotometer (GE Healthcare, Uppsala, Sweden).

2.2.2 Formation of the second disulfide bridge: Iodine mediated simultaneous deprotection/oxidation

Purified Cys(Acm)-containing peptides with one disulfide bridge was dissolved at a concentration of 0.6 mM in a mixed solvent consisting of 10% (v/v) ACN and 80% (v/v) acetic acid. Solid iodine (5 equivalent per Acm) and HCl (1.5 equivalent per Acm) were then added to the peptide solution. The reaction was allowed to proceed with vigorous stirring for 1 hr before quenching with a 1 M ascorbic acid solution drop-wise until a colorless solution was obtained. The reaction mixture was diluted 4-fold before loading into the Jupiter Proteo, 4 μ , 90 Å (15 × 250 mm) column for purification using the segmented gradient elution method described in Section 2.1.3. The purified peptides were verified using ESI-MS (i.e. mass reduction of 144 Da) before lyophilization.

2.3 Oxidative Folding of Fully Reduced Peptides

2.3.1 Air oxidation

The buffer used for air oxidation was 0.1 M Tris-HCl, pH 8.0, containing 10% (v/v) ACN. Fully reduced peptides (all cysteine residues derived from Cys(Trt) derivative) was dissolved at a concentration of 0.1 mM. The solution was stirred in an open atmosphere, and the progress of the reaction was monitored using the Ellman's test (Section 2.2.1.2). When the reaction was completed (negative Ellman's test), the pH of the solution was adjusted to pH 2 using concentrated HCl. For air oxidation in the presence of denaturant, 6 M guanidine hydrochloride (Gn.HCl) was included in the buffer.

2.3.2 Oxidation in the presence of redox reagents

The buffer used for redox reagent-mediated oxidation was 0.1 M Tris-HCl, pH 8.0, containing 1 mM EDTA, 2 mM reduced glutathione, 1 mM oxidized glutathione and 10% (v/v) ACN. Fully reduced peptides (all cysteine residues derived from Cys(Trt) derivative) was dissolved at a concentration of 0.1 mM. The solution was then purged with nitrogen gas for 5 mins before the reaction tube was sealed. The reaction was allowed to proceed with vigorous stirring for 48 hrs before the pH of the solution was adjusted to pH 2 with concentrated HCl. For redox reagent-mediated oxidation in the presence of denaturant or high salt content, 6 M Gn.HCl or 0.5 M NaCl was included in the buffer, respectively.

2.4 Chromatographic Separation of Structural Isoforms Obtained from Oxidative Folding Studies

2.4.1 Structural isoforms of t-TM EGF D4

Structural isoforms of t-TM EGF D4 obtained from oxidative folding (Section 2.3) were separated using reversed-phase HPLC with a Cosmosil Cholesterol, 5 μ , 120 Å (4.6 \times 250 mm) column (Nacalai Tesque, Kyoto, Japan). A segmented gradient elution method involving TFA as the counter-ion (constant concentration of 0.1% v/v), and MeOH as the organic modifier (maximum 60% v/v) was used.

2.4.2 Structural isoforms of t-TM EGF D5

Structural isoforms of t-TM EGF D5 obtained from oxidative folding (Section 2.3) were separated using reversed-phase HPLC with a Kinetex™ PFP, 2.6 μ , 100 Å (4.6 \times 100 mm) column (Phenomenex, Torrance, California, USA). A segmented gradient elution method involving heptafluorobutyric acid (HFBA) as the counter-ion (constant concentration of 10 mM), and MeOH as the organic modifier (maximum 80% v/v) was used.

2.4.3 Structural isoforms of t-TM EGF D4 (Y25T)

Structural isoforms of t-TM EGF D4 (Y25T) obtained from oxidative folding (Section 2.3) were separated using reversed-phase HPLC with a Cosmosil Cholesterol, 5 μ , 120 Å (4.6 \times 250 mm) column (Nacalai Tesque, Kyoto, Japan). A segmented gradient elution method involving HFBA as the counter-ion

(constant concentration of 10 mM), and MeOH as the organic modifier (maximum 80% v/v) was used.

2.4.4 Calculation of peak area

The amount of each structural isoform obtained is correlated to the area of its corresponding peak in the chromatogram. The peak area was calculated using the “peak integration” function of the UNICORN protein purification software (GE Healthcare, Uppsala, Sweden). Skim procedures were applied when deemed necessary to improve accuracy of calculations.

2.4.5 Statistical analysis

All oxidative folding experiments were performed in triplicates. The amount of each structural isoform obtained from each replicate was expressed as percentage values before the average and standard deviation values were calculated.

The Student’s t-test (independent samples) (Eq. 1 to Eq. 5, Table 2.1) was used to test for significant differences in the proportion of corresponding structural isoforms obtained from two different oxidative folding conditions. It should be noted that for a parametric test, the direct input of percentage data is not recommended. Thus, in accordance to a solution recommended by Zar [82], an arcsine transformation (Eq. 6, Table 2.1) was performed on all percentage values (from each replicate) before the statistical test was performed.

Student's t-test (Independent samples)

$$\text{Mean difference = (Samples)} = (\bar{x}_2 - \bar{x}_1) \quad (\text{Eq. 1})$$

$$\text{Mean difference = (Null hypothesis)} = (\mu_2 - \mu_1) \quad (\text{Eq. 2})$$

$$\text{Standard error of difference = (Samples)} = \sqrt{\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)} \quad (\text{Eq. 3})$$

$$\text{Test statistic = } t^* = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{\sqrt{\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)}} \quad (\text{Eq. 4})$$

$$\text{Degrees of freedom} = df = n_1 + n_2 - 2 \quad (\text{Eq. 5})$$

Arcsine transformation (degrees)

$$\text{Arcsine transformation} = \sin^{-1} \sqrt{\frac{x}{100}} \times \frac{180}{\pi} \quad (\text{Eq. 6})$$

Table 2.1 Annotation for statistical formulas (Eq. 1 to Eq. 6)

Student's t-test (Independent samples)	
\bar{x}_1	mean of arcsine transformed values — oxidative folding condition 1
\bar{x}_2	mean of arcsine transformed values — oxidative folding conditions 2
$\mu_2 - \mu_1$	H_0 : no difference = 0

s_1	standard deviation of arcsine transformed values — oxidative folding condition 1
s_2	standard deviation of arcsine transformed values — oxidative folding condition 2
n_1	number of trials for oxidative folding condition 1
n_2	number of trials for oxidative folding condition 2
Arcsine transformation	
x	percentage value

The p-values for the Pearson’s chi-square test were obtained using the “p-value calculator for the chi-square test” — <http://www.danielsoper.com/statcalc/calc11.aspx>

The p-values for the Student’s t-test (independent samples) were obtained using the “p-value calculator for the Student t-test” — <http://www.danielsoper.com/statcalc/calc08.aspx>



Chapter 3: Results and Discussion



3.1 Synthesis of Truncated TM EGF D4 and TM EGF D5 Structural Isoforms

To identify structural determinants in the 45- and 39-residue long domains of TM EGF D4 and TM EGF D5, respectively, it is important to simplify the search by narrowing down the region in which the determinants are located. Since the structural/disulfide-connectivity difference of TM EGF D4 and D5 are restricted to the first two disulfide bonds within their N-terminal segments (encompassing C₁ to C₄) (Figure 3.1), it was of interest to determine whether the structural determinants of each domain are located locally within that segment or if the C-terminal segment of the domain (encompassing C₅ to C₆) had a role in influencing the different disulfide-connectivity preference of the front segment.

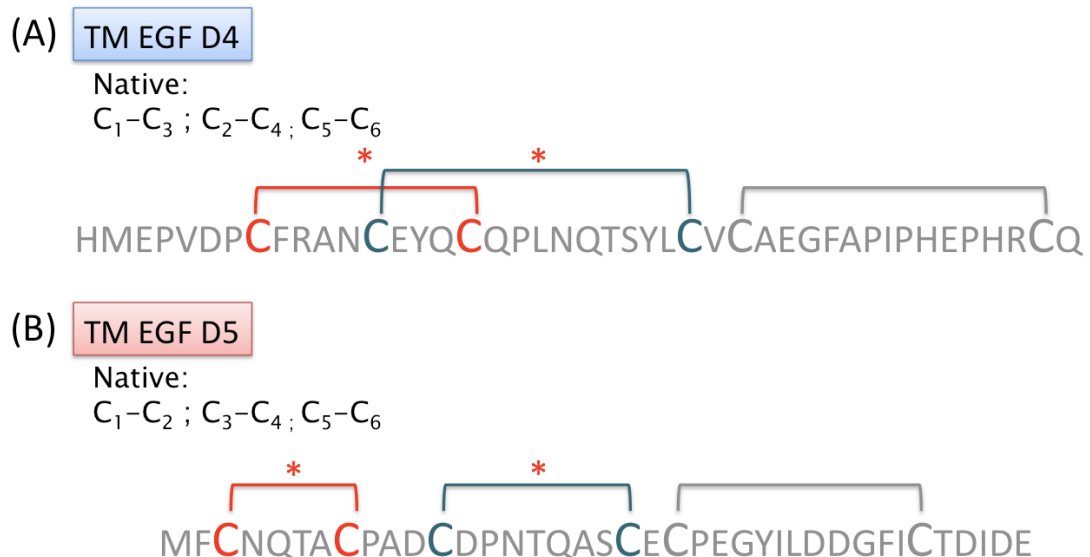


Figure 3.1 A comparison between the disulfide-connectivity of (A) TM EGF D4 and (B) TM EGF D5. The difference in disulfide-connectivity lies in the first two disulfide bonds of the respective domains, as denoted by asterisks (*). The disulfide-connectivity of C₁ to C₄ is C₁-C₃, C₂-C₄ for TM EGF D4, while it is C₁-C₂, C₃-C₄ for TM EGF D5.

To do so, truncated version of TM EGF D4 and D5 (t-TM EGF D4 and t-TM EGF D5) lacking the segment encompassing C₅ to C₆ were used for *in vitro* oxidative folding studies. The structural isoforms obtained from these studies were identified by comparing and matching their retention volume with those of regioselectively-synthesized structural isoforms.

These regioselectively-synthesized structural isoforms were generated using an orthogonal cysteine-protection scheme depicted in Figure 3.2 (details in Materials and Methods, Section 2.1 and Section 2.2). The retention volume and elution order of each individual structural isoform was then determined by reversed-phase HPLC on a Cosmosil Cholester column for t-TM EGF D4 isoforms, and a Kinetex PFP column for t-TM EGF D5 isoforms (details in Section 2.4).

3.1.1 Elution characteristics of t-TM EGF D4 structural isoforms

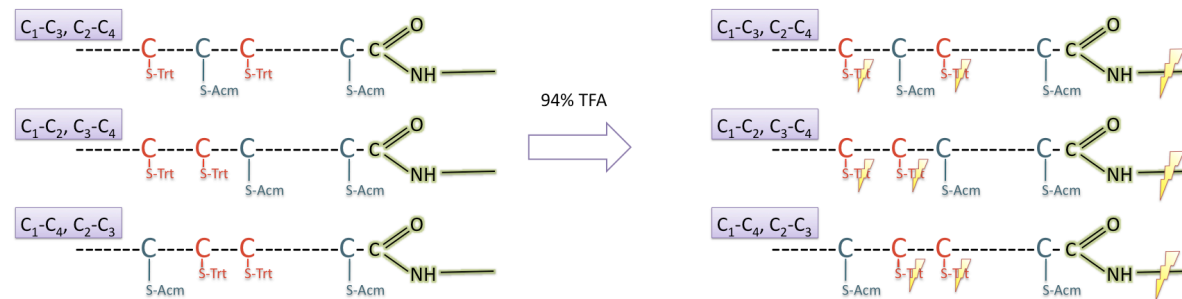
Human t-TM EGF D4 has a sequence of “HMEPVDPFCFRANCEYQCQPLNQT SYLCV”. The regioselective synthesis of t-TM EGF D4 structural isoforms were successful as the observed average mass of each isoform correspond well with the theoretical (fully oxidized) average mass of 3284.7 Da (Table 3.1 and Appendix (AP) Figure A3.1 to A3.3):

Table 3.1 *Observed versus theoretical mass of t-TM EGF D4 structural isoforms*

Structural Isoform	Observed Average Mass (Da) ^a	Theoretical Average Mass (Da)
C ₁ -C ₃ , C ₂ -C ₄ (Native)	3284.69	3284.7
C ₁ -C ₂ , C ₃ -C ₄	3284.68	
C ₁ -C ₄ , C ₂ -C ₃	3284.72	

^a Observed masses were calculated from ESI-MS data obtained from Perkin-Elmer Sciex API 300 LC/MS/MS system using the “Peptide Reconstruct” function of the Analyst[®] software.

Stage 1: Peptide synthesis and cleavage



Stage 2: Regioselective folding

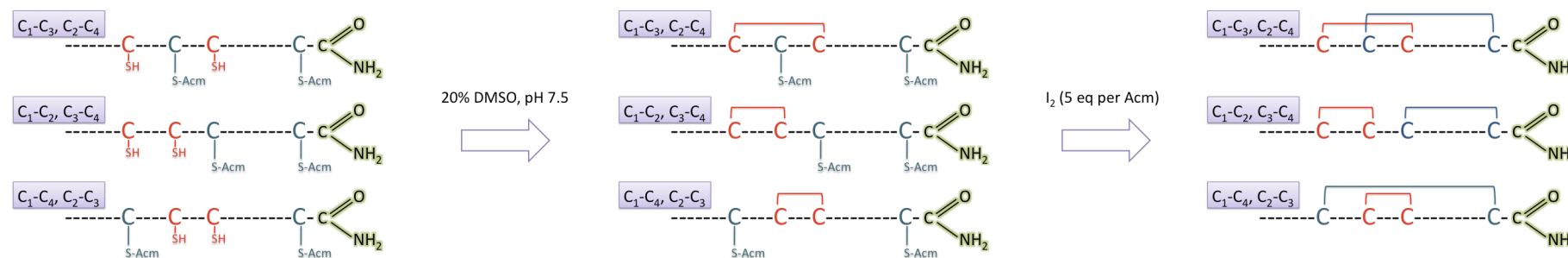


Figure 3.2 Regioselective synthesis of t-TM EGF D4 and t-TM EGF D5. In stage 1, the peptide chain was assembled by solid phase peptide synthesis on a rink amide-based resin. Cysteine residues with S-trityl (Trt) or S-acetamidomethyl (Acm) side-chain protection groups were incorporated at specific locations along the peptide chain. Upon treatment with high concentration of TFA, the peptide chain was released from the solid phase with simultaneous removal of all side-chain protection groups except Acm-groups on cysteine residues. In stage 2, the first disulfide bond was formed by DMSO-mediated oxidation of the two free cysteine residues previously derived from Cys(Trt). The second disulfide bond was formed by iodine treatment (at 5 equivalent per Acm) which mediated the simultaneous deprotection and oxidation of Cys(Acm) residues.

The elution order of t-TM EGF D4 structural isoforms based on reversed-phase analysis using the Cosmosil Cholesterol column is as follows: C₁-C₃, C₂-C₄ (native) (lowest retention volume), followed by C₁-C₂, C₃-C₄, and C₁-C₄, C₂-C₃ (highest retention volume).

As a gradual decrease in retention volume was observed with every use of the column (due to reasons yet unknown), each structural isoforms was re-analyzed individually on the column with every set of oxidative folding experiments to re-establish the retention volume of each structural isoform. This is to ensure the validity of the retention volume-based identification of various structural isoforms even though the elution order of the isoforms was not affected (i.e. retention volume of each isoform decrease by the same factor, therefore, not affecting the order of elution).

3.1.2 Elution characteristics of t-TM EGF D5 structural isoforms

Human t-TM EGF D5 has a sequence of “MFCNQ TACPADCDPNTQASCE”. The regioselective synthesis of t-TM EGF D5 structural isoforms were successful as the observed average mass of each isoform correspond well with the theoretical (fully oxidized) average mass of 2244.5 Da (Table 3.2 and AP Figure A3.4 to A3.6):

Table 3.2 Observed versus theoretical mass of t-TM EGF D5 structural isoforms

Structural Isoform	Observed Average Mass (Da) ^a	Theoretical Average Mass (Da)
C ₁ -C ₃ , C ₂ -C ₄	2244.49	2244.5
C ₁ -C ₂ , C ₃ -C ₄ (Native)	2244.45	
C ₁ -C ₄ , C ₂ -C ₃	2244.51	

^a Observed masses were calculated from ESI-MS data obtained from Perkin-Elmer Sciex API 300 LC/MS/MS system using the “Peptide Reconstruct” function of the Analyst[®] software.

The elution order of t-TM EGF D5 structural isoforms based on reversed-phase analysis using the Kinetex PFP column is as follows: C₁-C₂, C₃-C₄ (native) (lowest retention volume), followed by C₁-C₃, C₂-C₄, and C₁-C₄, C₂-C₃ (highest retention volume).

3.2 The *in vitro* Folding Tendencies of t-TM EGF D4 and t-TM EGF D5

To determine the folding tendency of t-TM EGF D4 and t-TM EGF D5, fully reduced peptides with all free cysteine residues, were placed in high pH (pH 8.0) buffer for *in vitro* oxidative folding. Two oxidative conditions were used: (a) Air oxidation, which makes use of atmospheric oxygen. Here, the oxidation process going through a series of free radical intermediates [40] and often results in cleaner products. However, the dominant product obtained may not represent the most thermodynamically favorable conformation; (b) Redox system, which involves the use of reduced:oxidized glutathione at a ratio of 2:1. These compounds catalyze disulfide exchange reactions resulting in the most thermodynamically favorable status of the cysteine residues [83]. However, it has the disadvantage of generating additional products which corresponds to peptides containing intermolecular disulfides between the peptide and glutathione [84].

3.2.1 *In vitro* oxidative folding of t-TM EGF D4

Oxidative folding of reduced t-TM EGF D4 was performed using air oxidation and redox reagent-mediated folding:

- (a) Air oxidation-mediated folding of reduced t-TM EGF D4 was monitored by the Ellman's test and the reaction was deemed complete when a negative test result was obtained. For t-TM EGF D4, the completion of oxidative folding took approximately 98 hrs. Structural isoforms obtained from the

reaction were resolved by reversed-phase chromatography and as expected, three monomeric isoforms were obtained (Figure 3.3).

(b) Redox reagent-mediated folding of t-TM EGF D4 was allowed to proceed for 48 hrs before the reaction was quenched by acidification. Like air oxidation, three monomeric isoforms were obtained (Figure 3.4).

In both cases, the retention volume of the three monomeric isoforms obtained matched well with that of regioselectively-synthesized structural isoforms. This enabled the identification of peaks in the chromatogram, and the relative proportions of the three isoforms were calculated based on the area of their respective peaks (Table 3.3A: Purple columns).

Results from both oxidative folding experiments showed that t-TM EGF D4 has a folding preference towards the C₁-C₃, C₂-C₄ (native) isoform as it constitutes the highest percentage of t-TM EGF D4 structural isoforms obtained in both cases.

Pairwise comparison of corresponding structural isoforms from air oxidation and redox reagent-mediated oxidation studies using Student's t-test (with arcsine transformed-percentage values) showed significant differences (at 0.05 level of significance) in the proportions of C₁-C₃, C₂-C₄ (native) and C₁-C₂, C₃-C₄ structural isoforms obtained between the two studies. Here, an increase and decrease in the proportion of C₁-C₃, C₂-C₄ (native) and C₁-C₂, C₃-C₄ structural isoforms, respectively, were observed when t-TM EGF D4 was folded via the redox buffer system (Figure 3.5 and AP Table A3.1).

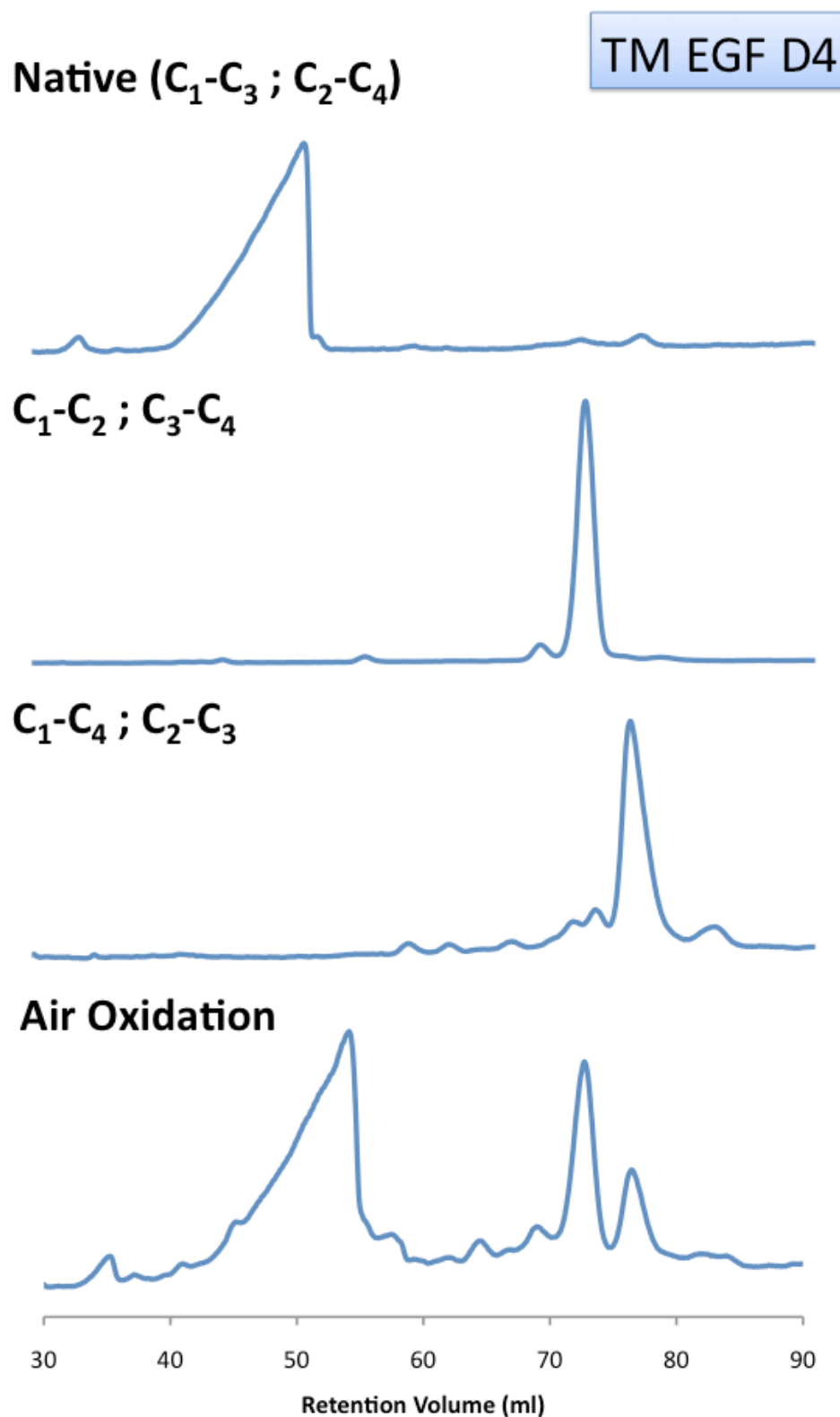


Figure 3.3 Analysis of t-TM EGF D4 air oxidation products by reversed-phase chromatography. Retention volume of the three monomeric structural isoforms obtained were compared with that of regioselectively-synthesized structural isoforms.

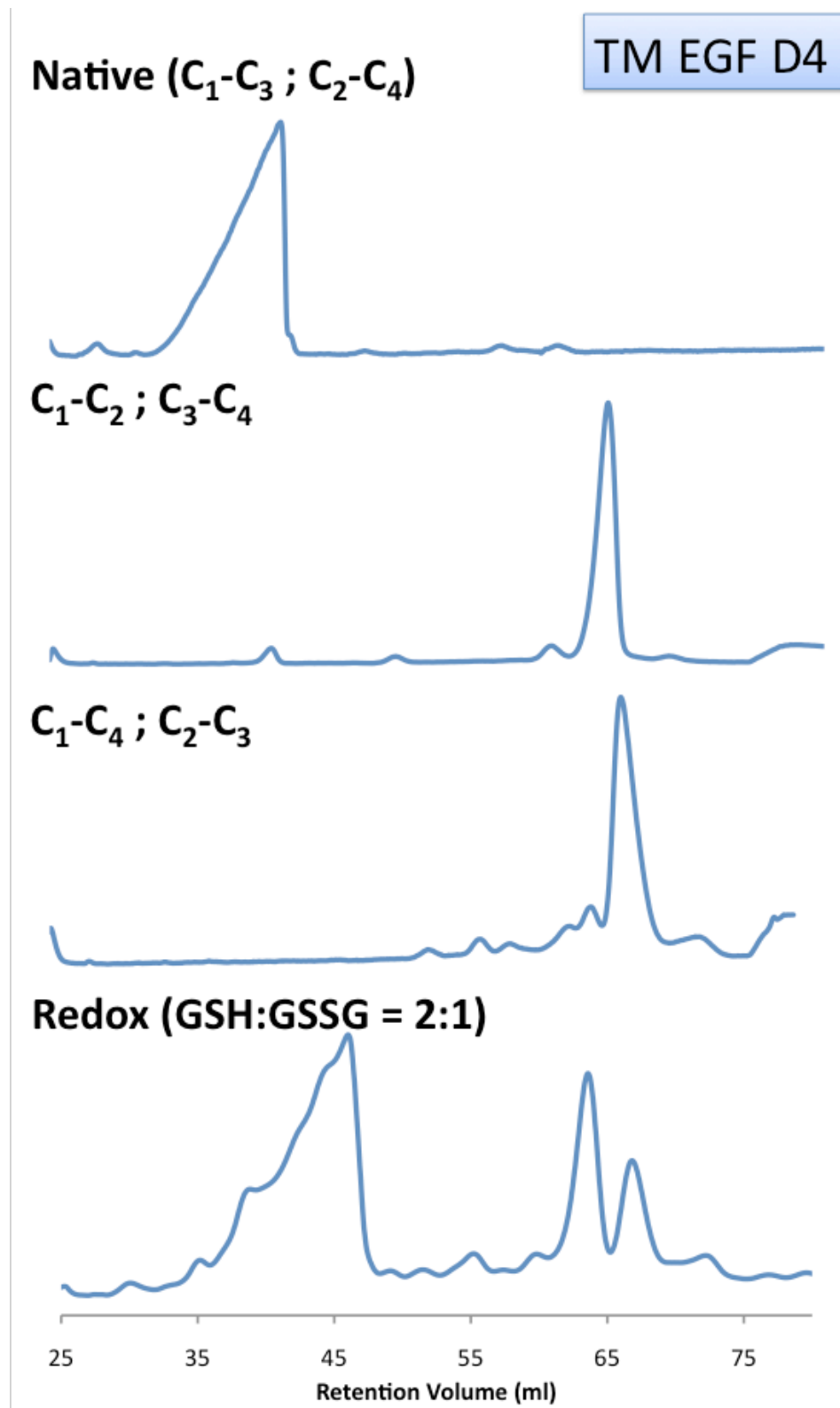


Figure 3.4 Analysis of t-TM EGF D4 redox reagent-mediated oxidation products by reversed-phase chromatography. Retention volume of the three monomeric structural isoforms obtained were compared with that of regioselectively-synthesized structural isoforms. Abbreviations: GSH — Reduced glutathione; GSSG — Oxidized glutathione.

Table 3.3 Percentages of structural isoforms obtained from oxidative folding of t-TM EGF D4 and t-TM EGF D5 in various conditions

(A) Oxidative folding of t-TM EGF D4					
Structural Isoforms	(-- ^{a, b})		+6 M Gn.HCl ^c		+0.5 M NaCl ^d
	Air Oxidation (%)	Redox Buffer System (%)	Air Oxidation (%)	Redox Buffer System (%)	Redox Buffer System (%)
C ₁ -C ₃ , C ₂ -C ₄ (Native)	67.53 ± 0.69	69.08 ± 0.57	18.48 ± 1.15	31.31 ± 0.98	74.66 ± 0.87
C ₁ -C ₂ , C ₃ -C ₄	21.13 ± 0.67	18.96 ± 0.57	52.72 ± 0.94	47.28 ± 0.32	17.18 ± 0.45
C ₁ -C ₄ , C ₂ -C ₃	11.34 ± 0.63	11.96 ± 0.27	28.80 ± 0.82	21.42 ± 0.71	8.17 ± 0.49

(B) Oxidative folding of t-TM EGF D5					
Structural Isoforms	(-- ^{a, b})		+6 M Gn.HCl ^c		+0.5 M NaCl ^d
	Air Oxidation (%)	Redox Buffer System (%)	Air Oxidation (%)	Redox Buffer System (%)	Redox Buffer System (%)
C ₁ -C ₃ , C ₂ -C ₄	20.57 ± 0.34	20.36 ± 0.11	17.69 ± 0.35	19.11 ± 0.15	23.32 ± 0.43
C ₁ -C ₂ , C ₃ -C ₄ (Native)	60.40 ± 0.64	60.67 ± 1.07	61.80 ± 0.91	61.46 ± 0.77	54.40 ± 1.78
C ₁ -C ₄ , C ₂ -C ₃	19.03 ± 0.30	18.97 ± 0.98	20.52 ± 0.67	19.43 ± 0.63	22.28 ± 1.98

^a Discussion in Section 3.2, ^b Normal oxidative folding conditions (without denaturant or salt)

^c Discussion in Section 3.3

^d Discussion in Section 3.4

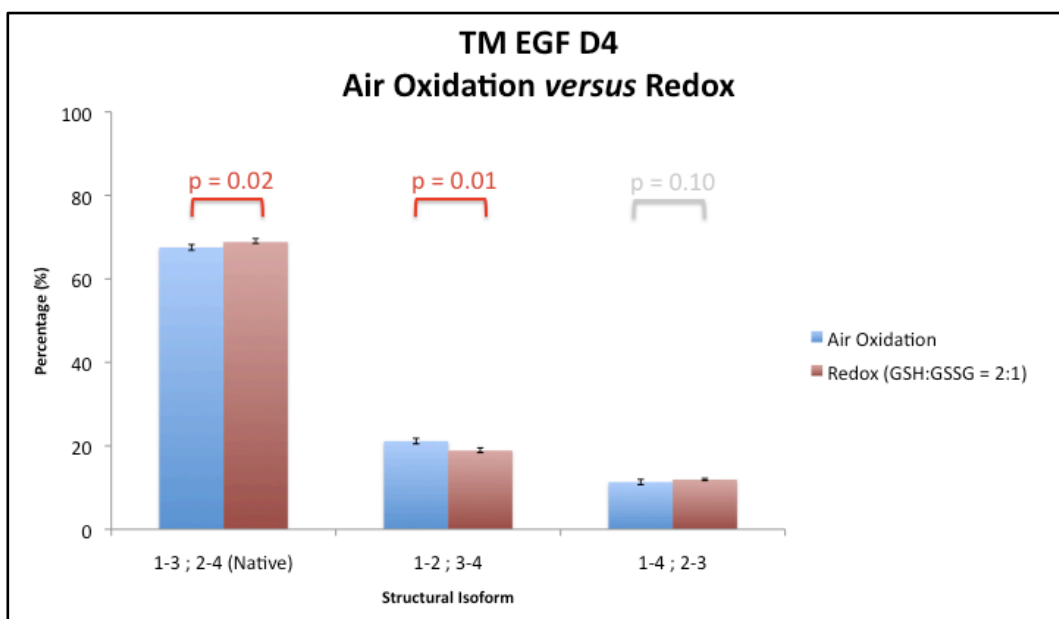


Figure 3.5 Pairwise comparison of t-TM EGF D4 structural isoform proportions obtained from air oxidation and redox reagent-mediated oxidation studies. Student's t-test (independent samples) using arcsine transformed-values were used for the calculation of probability (p)-values. Difference in proportion between corresponding structural isoforms is deemed to be significant when the p-value is less than 0.05 (one-tailed).

3.2.2 *In vitro* oxidative folding of t-TM EGF D5

Oxidative folding of reduced t-TM EGF D5 was performed using air oxidation and redox reagent-mediated folding:

- (a) Air oxidation-mediated folding of reduced t-TM EGF D5 was completed in approximately 72 hrs as judged by the Ellman's test. Structural isoforms obtained from the reaction were resolved by reversed-phase chromatography and as expected, three monomeric isoforms were obtained (Figure 3.6).
- (b) Folding of t-TM EGF D5 in redox buffer system was performed over 48 hrs. Like air oxidation, three monomeric isoforms were obtained (Figure 3.7).

Like t-TM EGF D4, the retention volume of t-TM EGF D5 structural isoforms

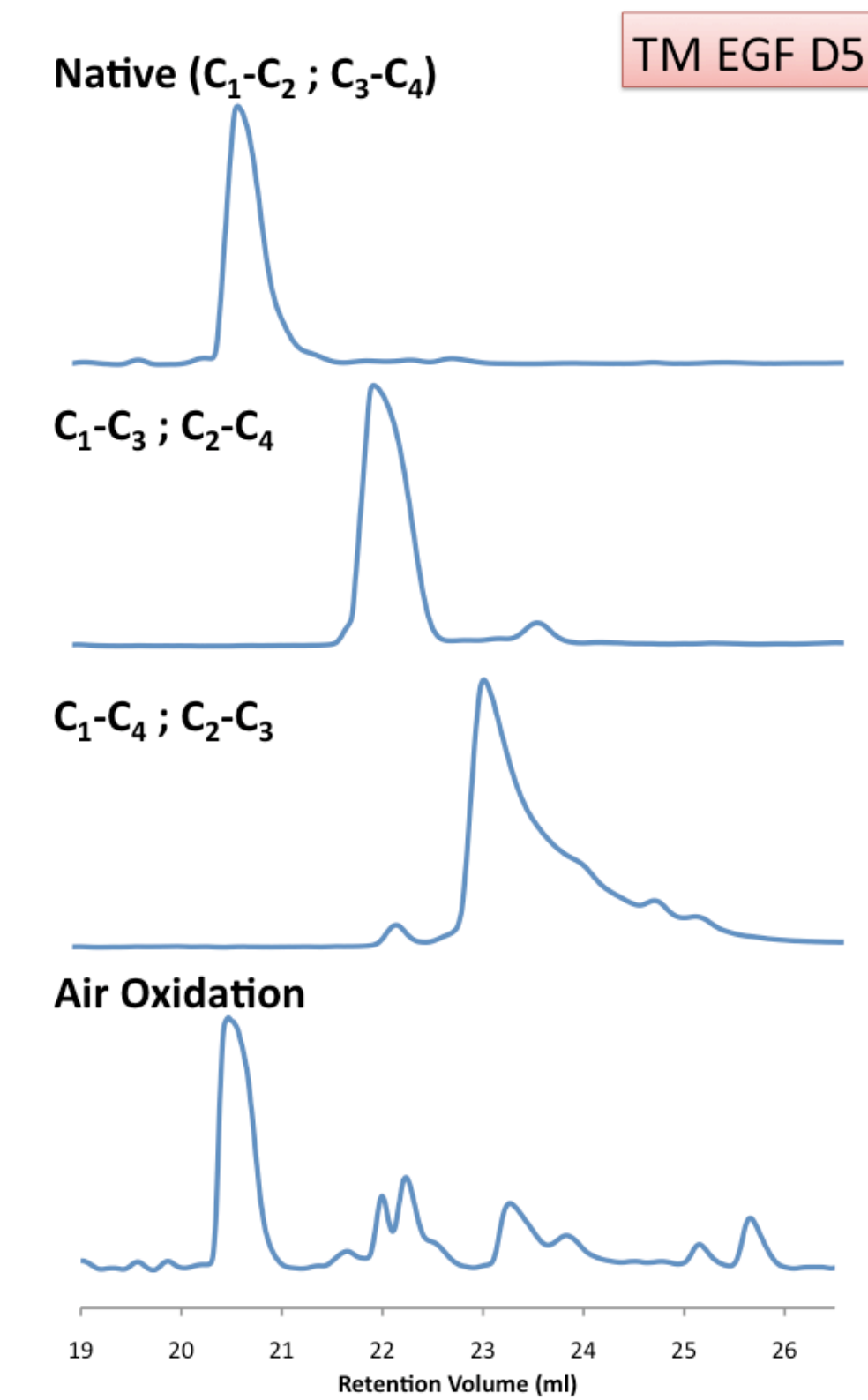


Figure 3.6 Analysis of t-TM EGF D5 air oxidation products by reversed-phase chromatography. Retention volume of the three monomeric structural isoforms obtained were compared with that of regioselectively-synthesized structural isoforms.

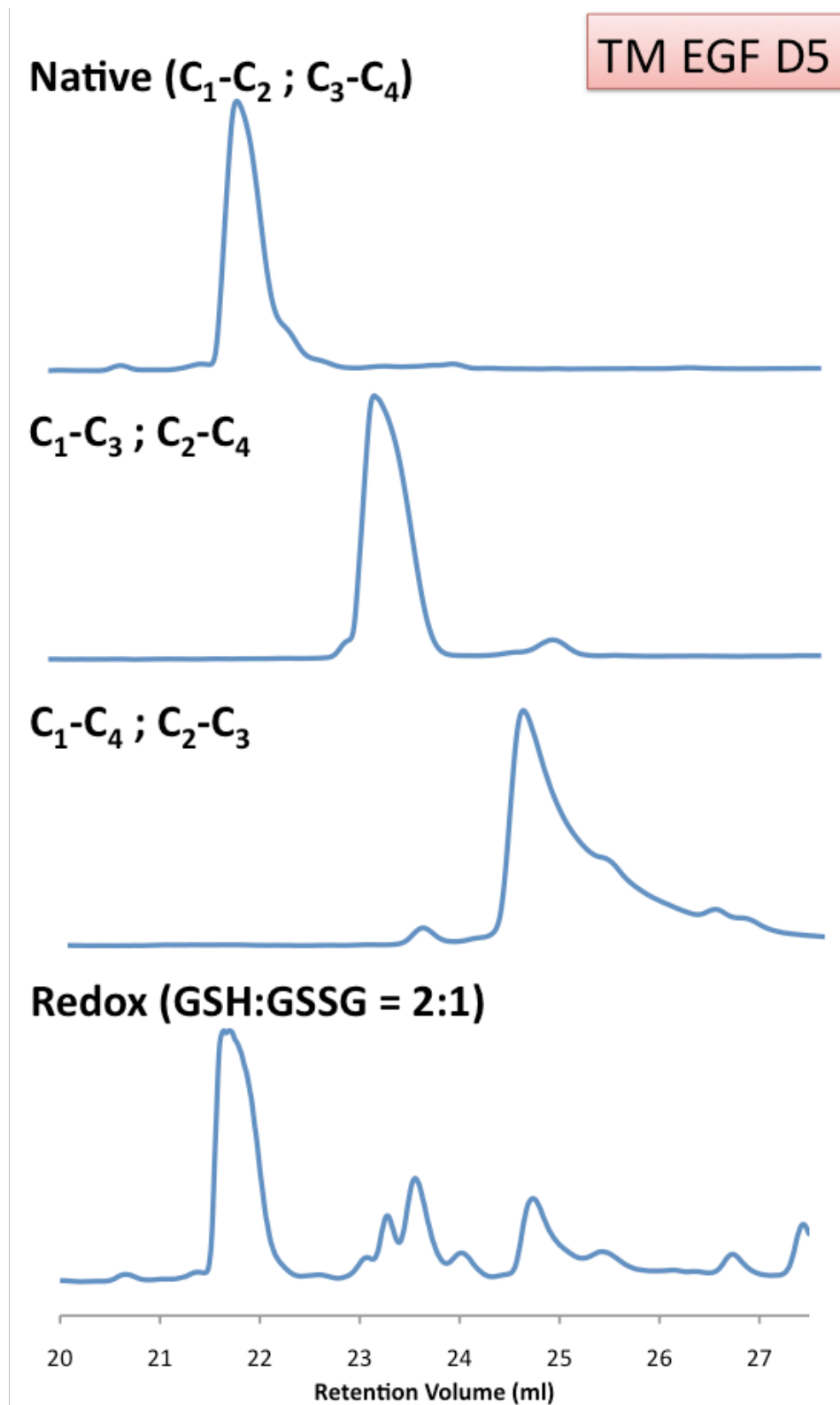


Figure 3.7 Analysis of t-TM EGF D5 redox reagent-mediated oxidation products by reversed-phase chromatography. Retention volume of the three monomeric structural isoforms obtained were compared with that of regioselectively-synthesized structural isoforms. Abbreviations: GSH — Reduced glutathione; GSSG — Oxidized glutathione.

obtained from both oxidative folding studies matched well with that of regioselectively-synthesized structural isoforms. Thus, peak identities were assigned and the relative proportions of the three isoforms were calculated based on the area of the respective peaks (Table 3.3B: Purple columns).

Unlike the folding tendency of t-TM EGF D4, results from these oxidative folding experiments showed that t-TM EGF D5 has a folding preference towards the C₁-C₂, C₃-C₄ isoform instead of the C₁-C₃, C₂-C₄ isoform. However, it should be noted that C₁-C₂, C₃-C₄ is the native disulfide-connectivity of t-TM EGF D5. Thus, in a way similar to t-TM EGF D4, t-TM EGF D5 has a folding preference towards its native isoform.

With regards to the difference in proportions of corresponding structural isoforms obtained from air oxidation and redox reagent-mediated oxidation studies, the Student's t-test (with arcsine transformed values) revealed no significant difference (at 0.05 level of significance) in all structural isoforms between the two studies (Figure 3.8 and AP Table A3.2).

3.2.3 Truncated TM EGF D4 and TM EGF D5 preferentially fold into their respective native isoform

Similar to air oxidation-mediated folding, the highest yield obtained from redox reagent-mediated folding was the respective native isoforms of both domains. This is approximately 70% for t-TM EGF D4 (C₁-C₃, C₂-C₄) and 60% for t-TM EGF D5 (C₁-C₂, C₃-C₄). From these redox-based experiments, it is reasonable to conclude that the respective native isoforms of both domains are the most thermodynamically stable among the three possible structural isoforms.

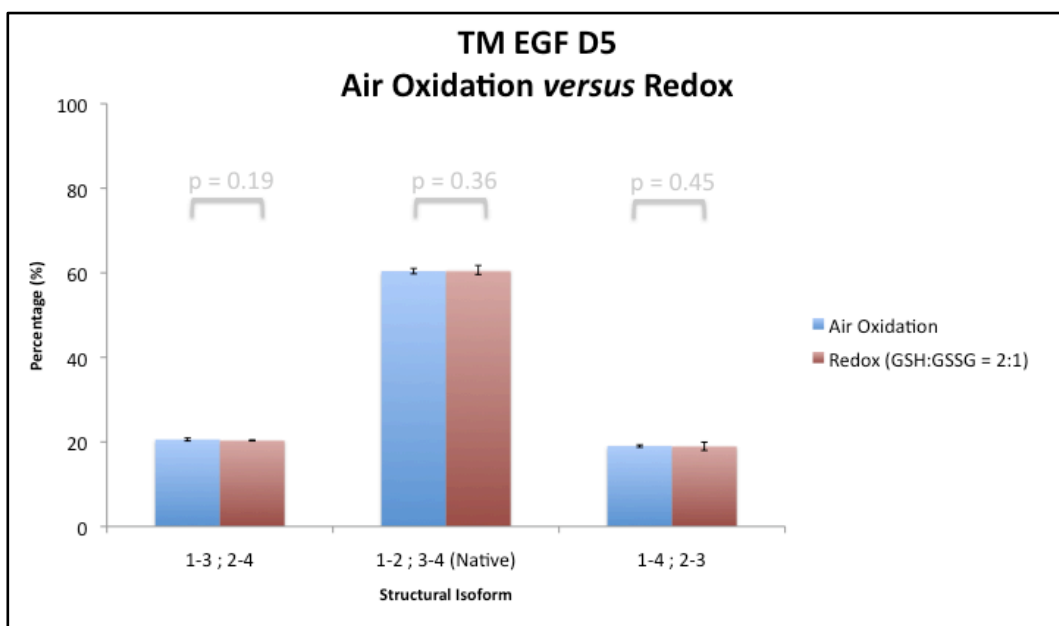


Figure 3.8 Pairwise comparison of t-TM EGF D5 structural isoform proportions obtained from air oxidation and redox reagent-mediated oxidation studies. Student's t-test (independent samples) using arcsine transformed-values were used for the calculation of probability (p)-values. Difference in proportion between corresponding structural isoforms is deemed to be significant when the p-value is less than 0.05 (one-tailed).

Together, these results demonstrated that fully reduced t-TM EGF D4 and t-TM EGF D5 still preferentially fold into their native, most thermodynamically stable, structural isoform even when the C-terminal segment of both EGF-like domains (encompassing C₅ and C₆) were absent. This prominent folding tendency suggests the existence of structural determinants that lies within the N-terminal segment (encompassing C₁ to C₄) of both domains and that their respective C-terminal segments do not play a major role in dictating the disulfide-connectivity of the first two disulfide bonds. Logically, the respective structural determinants of both domains must be of different or opposing properties so as to dictate the canonical *versus* non-canonical EGF-like domain fold of TM EGF D4 and TM EGF D5, respectively.

3.3 Contribution of Side-chain Interactions in the Folding Tendencies of t-TM EGF D4 and t-TM EGF D5

To identify the structural determinants which are located in the N-terminal segment of the EGF-like domain, it is important to identify the dominant forces that drives and stabilizes the fold of t-TM EGF D4 and t-TM EGF D5. To this end, it is of interest to detect any difference in the folding tendency of both truncated EGF-like domains upon manipulation of the oxidative folding environment. This would tell us the relative contribution of specific side-chain interactions in stabilizing the native fold of the domain and thus aid in the identification of the dominant force. Any knowledge of the dominant force would then indicate the physical-chemical properties of the amino acid residues involved in the folding code of the EGF-like domain.

To determine the role of side-chain interactions in dictating the folding tendency of t-TM EGF D4 and t-TM EGF D5, 6 M Gn.HCl was included in the oxidative folding buffer to disrupt side-chain interactions in the peptide. The extent of change in the folding tendency of both domains would be compared to ascertain if differences in the necessity of side-chain interactions is the main contributor to the two EGF-like domains' different folding tendency.

3.3.1 *In vitro* oxidative folding of t-TM EGF D4 in the presence of 6 M Gn.HCl

Air oxidation and redox reagent-mediated folding of t-TM EGF D4 was performed with the inclusion of 6 M Gn.HCl in the folding buffer:

- (a) For air oxidation in the presence of 6 M Gn.HCl, the reaction was completed in approximately 72 hrs as judged by the Ellman's test. Analysis by reversed-phase chromatography revealed that all three monomeric isoforms were obtained (Figure 3.9).
- (b) The redox reagent-mediated oxidative folding of t-TM EGF D4 in the presence of 6 M Gn.HCl was allowed to proceed for 48 hours. It also yielded three monomeric isoforms which were resolved by reversed-phase chromatography (Figure 3.10).

Peak identities were assigned and their respective peak area revealed that the highest yield obtained from both oxidative folding studies was the C₁-C₂, C₃-C₄ isoform (Table 3.3A: Green columns). This meant that t-TM EGF D4 has a folding preference towards the C₁-C₂, C₃-C₄ isoform, instead of its native C₁-C₃, C₂-C₄ isoform, when folded in the presence of denaturant (*Note*: more detailed discussion in Section 3.3.3).

Pairwise comparison of corresponding structural isoforms from both oxidative folding studies was performed. The Student's t-test (with arcsine transformed-percentage values) showed significant difference in the proportions of all structural isoforms obtained — i.e. When folding was performed via the redox buffer system in the presence of denaturant, the proportion of the C₁-C₃, C₂-C₄ (native) isoform obtained was much higher. This resulted in the concomitant decrease in the proportions of the C₁-C₂, C₃-C₄ and C₁-C₄, C₂-C₃ isoforms (Figure 3.11 and AP Table A3.3). The increase in the C₁-C₃, C₂-C₄ (native) isoform in redox buffer system could be attributed to the redox

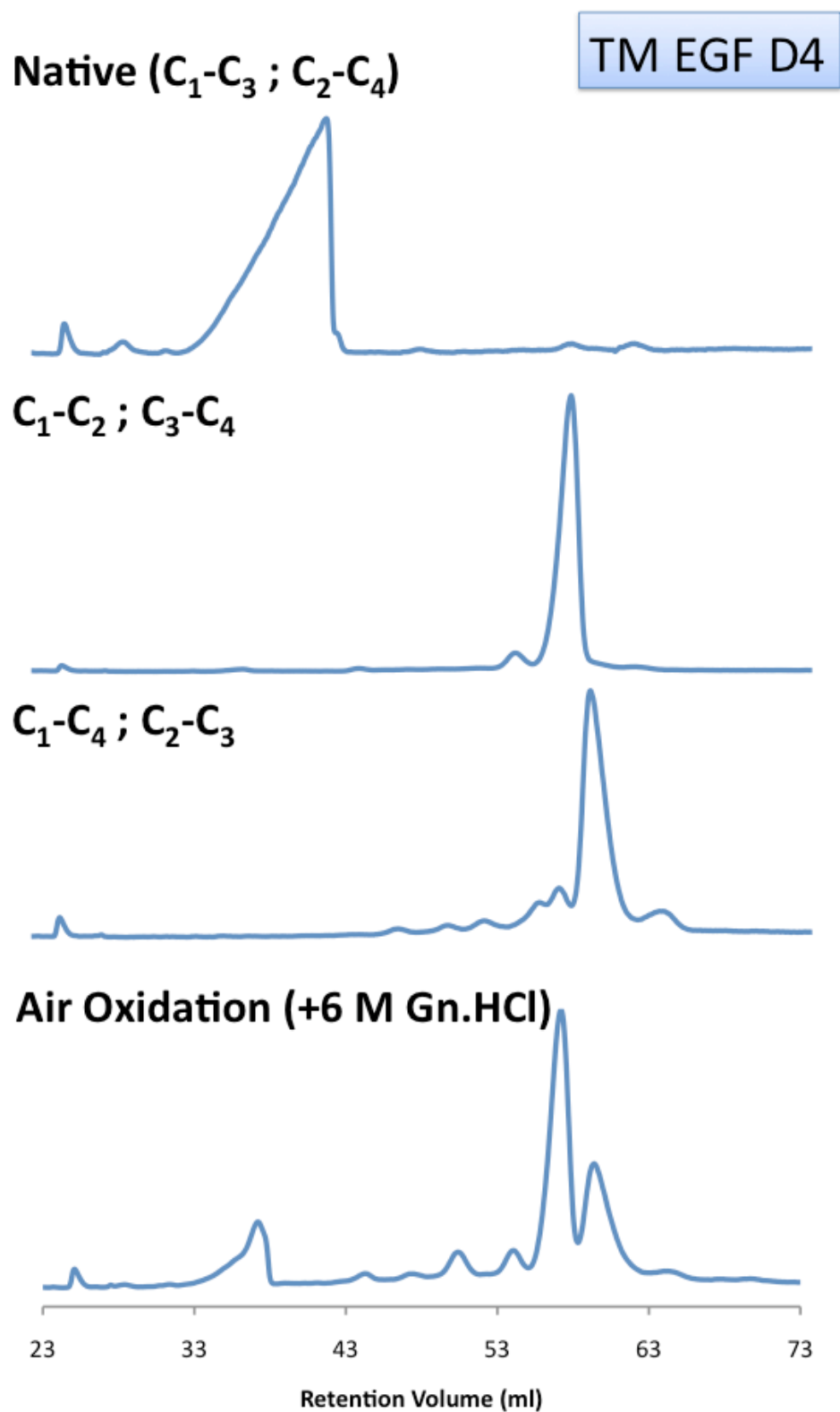


Figure 3.9 Analysis of t-TM EGF D4 products obtained from air oxidation in the presence of 6 M Gn.HCl by reversed-phase chromatography. Retention volume of the three monomeric structural isoforms obtained were compared with that of regioselectively-synthesized structural isoforms.

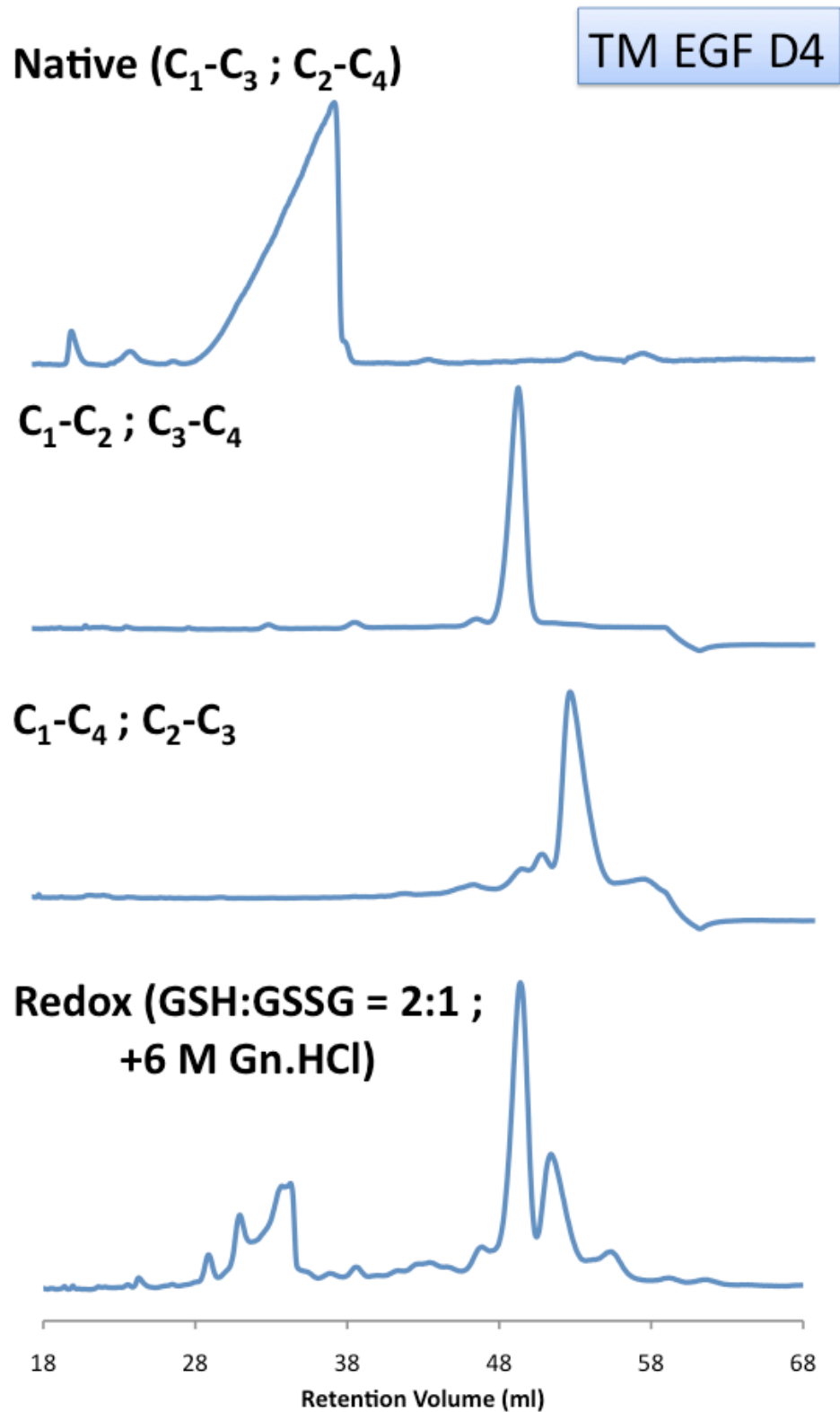


Figure 3.10 Analysis of t-TM EGF D4 products obtained from redox reagent-mediated oxidation in the presence of 6 M Gn.HCl by reversed-phase chromatography. Retention volume of the three monomeric structural isoforms obtained were compared with that of regiospecifically-synthesized structural isoforms. Abbreviations: GSH — Reduced glutathione; GSSG — Oxidized glutathione.

reagent-mediated increase in the most thermodynamically stable isoform of t-TM EGF D4 (i.e. C₁-C₃, C₂-C₄). Here, redox reagent-mediated oxidation had increase the proportion of C₁-C₃, C₂-C₄ isoform despite an overwhelming tendency for t-TM EGF D4 to fold into the C₁-C₂, C₃-C₄ isoform in the presence of 6 M Gn.HCl.

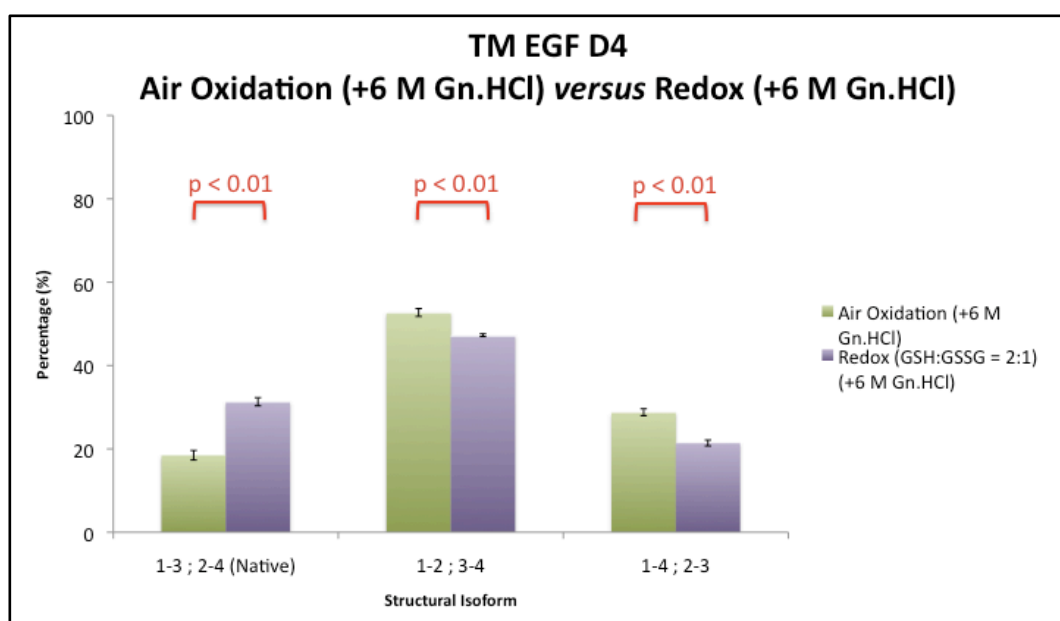


Figure 3.11 Pairwise comparison of t-TM EGF D4 structural isoform proportions obtained from air oxidation (with 6 M Gn.HCl) and redox reagent-mediated oxidation (with 6 M Gn.HCl) studies. Student's t-test (independent samples) using arcsine transformed-values were used for the calculation of probability (p)-values. Difference in proportion between corresponding structural isoforms is deemed to be significant when the p-value is less than 0.05 (one-tailed).

3.3.2 *In vitro* oxidative folding of t-TM EGF D5 in the presence of 6 M Gn.HCl

Air oxidation and redox reagent-mediated folding of t-TM EGF D5 was performed with the inclusion of 6 M Gn.HCl in the folding buffer:

- (a) In the presence of denaturant, air oxidation-mediated folding was completed in approximately 48 hrs as judged by the Ellman's test.

Subsequent analysis by reversed-phase chromatography showed that three structural isoforms were obtained (Figure 3.12).

- (b) The folding of t-TM EGF D5 using redox reagent-mediated oxidation in the presence of denaturant also yielded three structural isoforms (Figure 3.13).

Quantification of structural isoform proportions based on relative peak areas revealed that the folding tendency of t-TM EGF D5 was not affected by the presence of 6 M Gn.HCl in the oxidative folding buffers. t-TM EGF D5 still showed a folding preference towards its native (C₁-C₂, C₃-C₄) isoform (Table 3.3B: Green columns). This is in contrast to t-TM EGF D4 whose folding tendency was altered when denaturant was added to the oxidative folding buffers (*Note*: more detailed discussion in Section 3.3.3).

Pairwise comparison of corresponding structural isoforms from both set of experiments using Student's t-test (with arcsine transformed values) revealed no difference in proportions, except for a slightly larger percentage of C₁-C₃, C₂-C₄ isoform in the redox reagent mediated-experiments (Figure 3.14 and AP Table A3.4).

3.3.3 Side-chain interaction is necessary for the canonical C₁-C₃, C₂-C₄ fold of the EGF-like domain

Disruption of side-chain interactions using 6 M Gn.HCl in the oxidative folding buffer had a different effect on the folding tendencies of t-TM EGF D4 and t-TM EGF D5.

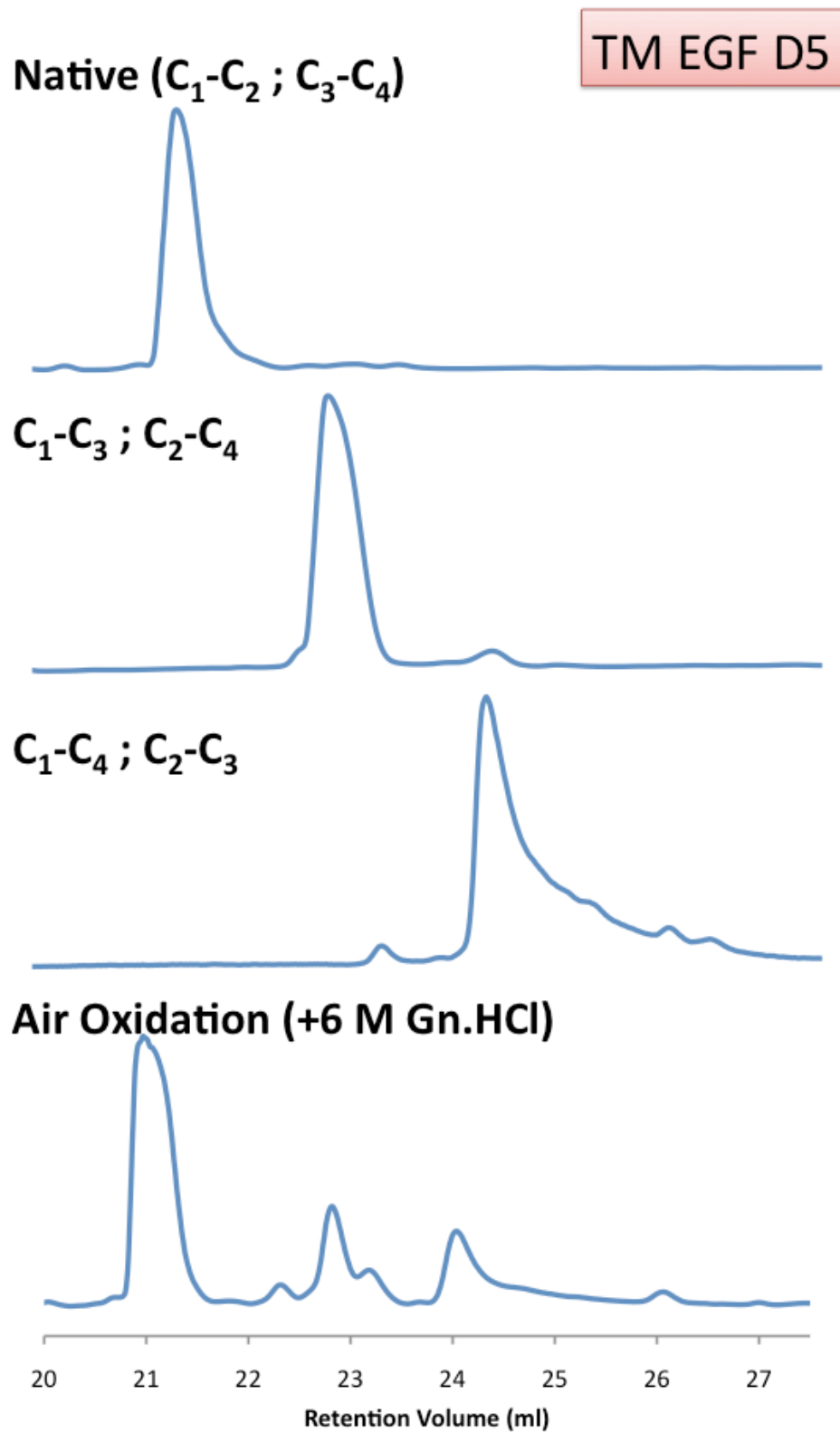


Figure 3.12 Analysis of t-TM EGF D5 products obtained from air oxidation in the presence of 6 M Gn.HCl by reversed-phase chromatography. Retention volume of the three monomeric structural isoforms obtained were compared with that of regiospecifically-synthesized structural isoforms.

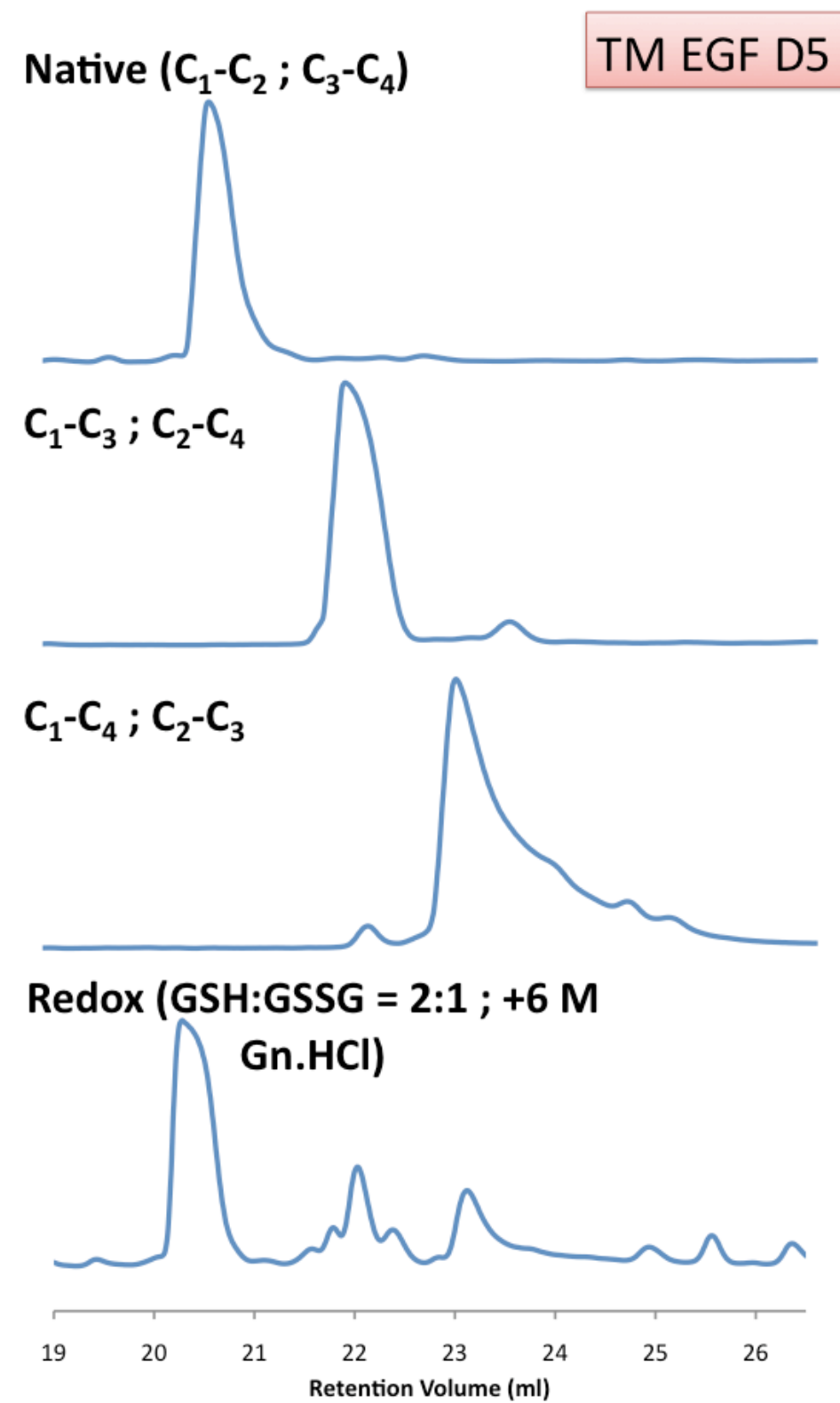


Figure 3.13 Analysis of t-TM EGF D5 products obtained from redox reagent-mediated oxidation in the presence of 6 M Gn.HCl by reversed-phase chromatography. Retention volume of the three monomeric structural isoforms obtained were compared with that of regioselectively-synthesized structural isoforms. Abbreviations: GSH — Reduced glutathione; GSSG — Oxidized glutathione.

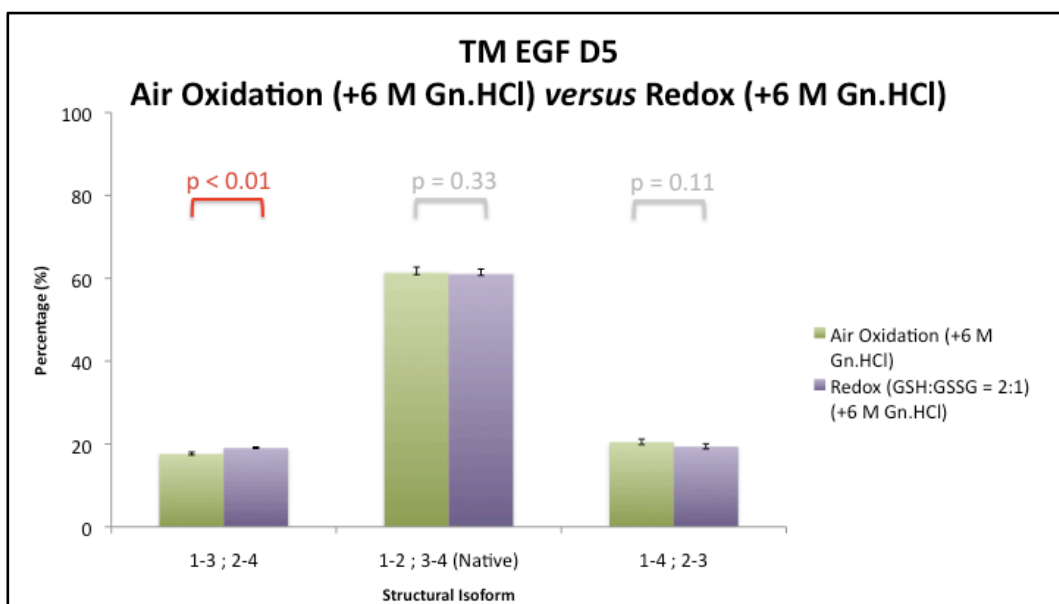


Figure 3.14 Pairwise comparison of t-TM EGF D5 structural isoform proportions obtained from air oxidation (with 6 M Gn.HCl) and redox reagent-mediated oxidation (with 6 M Gn.HCl) studies. Student's t-test (independent samples) using arcsine transformed-values were used for the calculation of probability (p)-values. Difference in proportion between corresponding structural isoforms is deemed to be significant when the p-value is less than 0.05 (one-tailed).

3.3.3.1 Disruption of side-chain interactions led to change in folding tendency of t-TM EGF D4

For t-TM EGF D4, the loss of side-chain interactions resulted in the change of folding tendency from that of C₁-C₃, C₂-C₄ (native) to C₁-C₂, C₃-C₄. This shift in folding tendency was rather prominent as the percentage of native isoform dropped from 67.53 ± 0.69% to 18.48 ± 1.15% when 6 M Gn.HCl was included in the air oxidation buffer, and from 69.08 ± 0.57% to 31.31 ± 0.97% when the denaturant was included in redox buffer. To put these numbers into perspective, the decreased native isoform proportion corresponds to only 0.26 (about one quarter) and 0.45 (about half) of the original air oxidation and redox reagent-mediated oxidation proportion, respectively.

For the C₁-C₂, C₃-C₄ isoform, the disruption of side-chain interactions in the folding peptide had benefited its numbers. When 6 M Gn.HCl was included in

the air oxidation buffer, the percentage of C₁-C₂, C₃-C₄ isoform increased from 21.13 ± 0.67% to 52.72 ± 0.94%. For the inclusion of denaturant into the redox buffer, the percentage of C₁-C₂, C₃-C₄ isoform increased from 18.96 ± 0.57% to 47.28 ± 0.31%. In both cases, the increased C₁-C₂, C₃-C₄ proportion corresponds to 2.5 times the original amount obtained from air oxidation and redox reagent-mediated oxidation.

To lend further support to the observation that the folding tendency of t-TM EGF D4 was affected when side-chain interactions were disrupted, the Student's t-test (with arcsine transformed values) was performed. Pairwise comparison of corresponding structural isoform proportions showed statistically significant decrease and increase in the proportions of C₁-C₃, C₂-C₄ and C₁-C₂, C₃-C₄ isoforms, respectively, when oxidative folding was conducted in the presence of denaturant (Figure 3.15 and Figure 3.16, AP Table A3.5 and AP Table A3.6).

Thus, the folding of t-TM EGF D4 into its native isoform is highly dependent upon the presence of side-chain interactions. In its absence, the t-TM EGF D4 opted for the fold with the C₁-C₂, C₃-C₄ disulfide-connectivity which is interestingly the native conformer of t-TM EGF D5, the non-canonical EGF-like domain.

3.3.3.2 Disruption of side-chain interactions did not affect the folding tendency of t-TM EGF D5

Based on the results presented in Section 3.3.2, it is now apparent that the folding tendency of t-TM EGF D5 was not affected by the disruption of side-

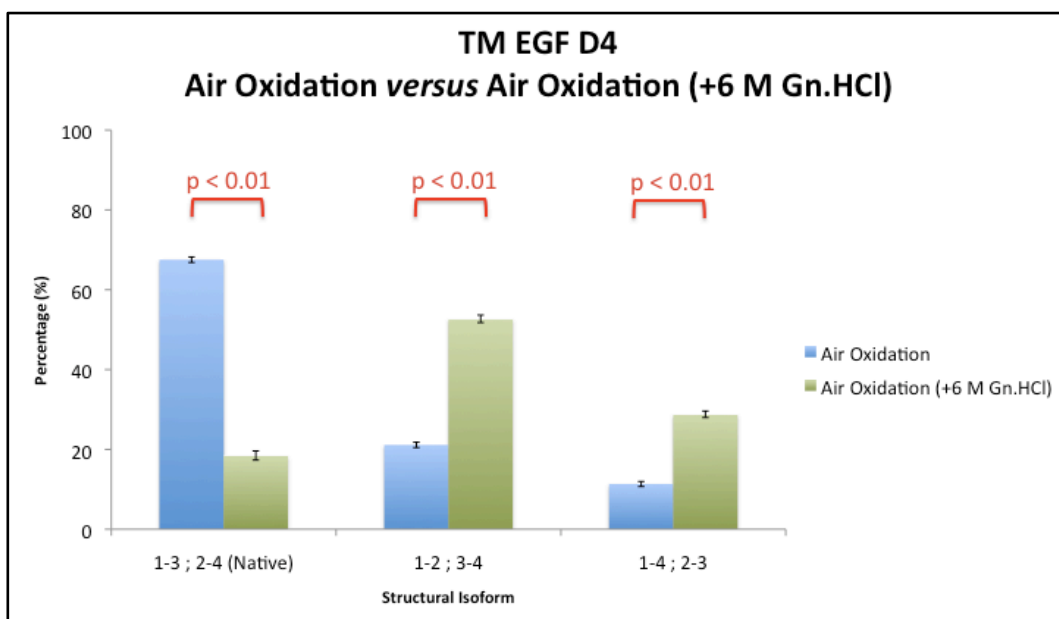


Figure 3.15 Pairwise comparison of t-TM EGF D4 structural isoform proportions obtained from air oxidation and air oxidation (with 6 M Gn.HCl) studies. Student's t-test (independent samples) using arcsine transformed-values were used for the calculation of probability (p)-values. Difference in proportion between corresponding structural isoforms is deemed to be significant when the p-value is less than 0.05 (one-tailed).

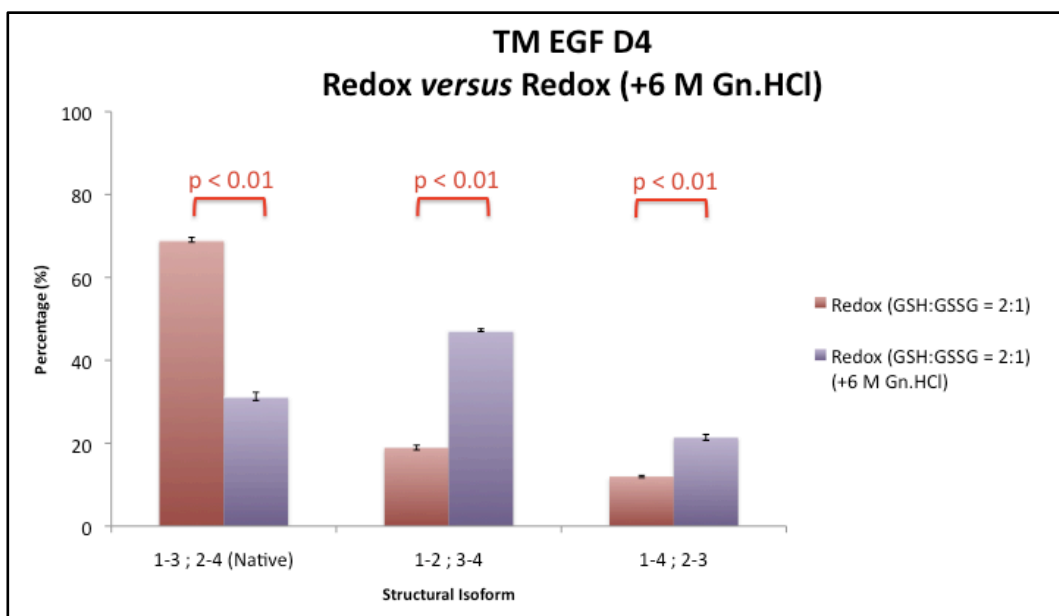


Figure 3.16 Pairwise comparison of t-TM EGF D4 structural isoform proportions obtained from redox reagent-mediated oxidation and redox reagent-mediated oxidation (with 6 M Gn.HCl) studies. Student's t-test (independent samples) using arcsine transformed-values were used for the calculation of probability (p)-values. Difference in proportion between corresponding structural isoforms is deemed to be significant when the p-value is less than 0.05 (one-tailed).

chain interactions. When looking at the exact numbers, the presence of 6 M Gn.HCl in both oxidative folding buffer had resulted in a slight but statistically significant decrease in the C₁-C₃, C₂-C₄ isoform, without affecting the proportion of the C₁-C₂, C₃-C₄ (native) isoform at all (Figure 3.17 and Figure 3.18, AP Table A3.7 and Table A3.8).

In conclusion, the results obtained from oxidative folding studies on t-TM EGF D4 and t-TM EGF D5 in the presence of 6 M Gn.HCl suggested that the absence of side-chain interactions generally decreases the C₁-C₃, C₂-C₄ isoform and increases the C₁-C₂, C₃-C₄ isoform regardless of the exact identities of the EGF-like domains (i.e. whether it is t-TM EGF D4 or t-TM EGF D5). In view of this generalized effect, the finding that the absence of side-chain interactions disfavors the C₁-C₃, C₂-C₄ conformer could be applicable to other canonical EGF-like domains as well.

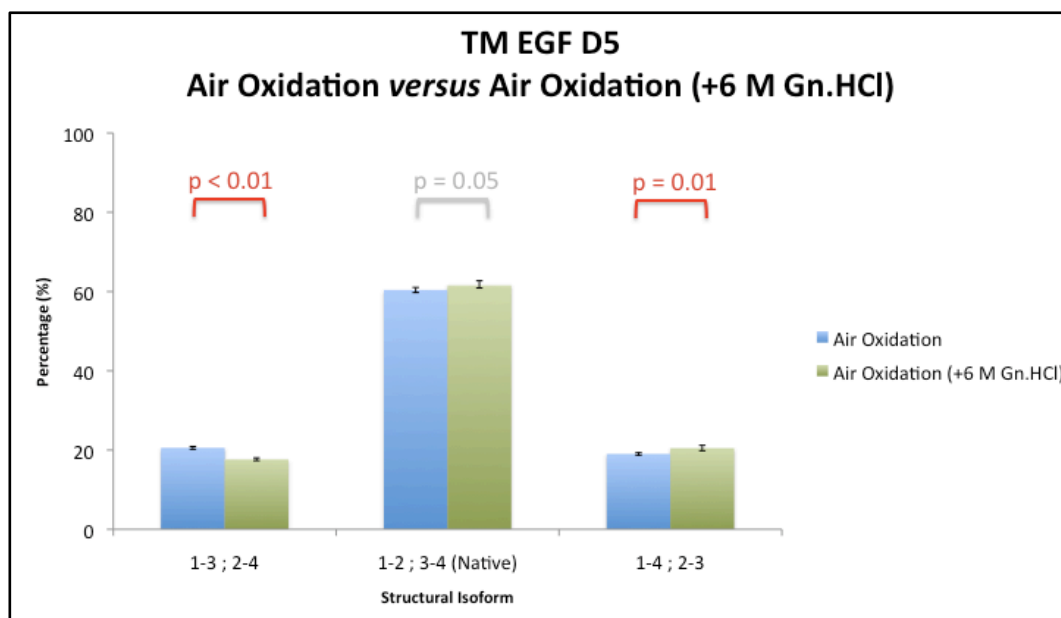


Figure 3.17 Pairwise comparison of t-TM EGF D5 structural isoform proportions obtained from air oxidation and air oxidation (with 6 M Gn.HCl) studies. Student's t-test (independent samples) using arcsine transformed-values were used for the calculation of probability (p)-values. Difference in proportion between corresponding structural isoforms is deemed to be significant when the p-value is less than 0.05 (one-tailed).

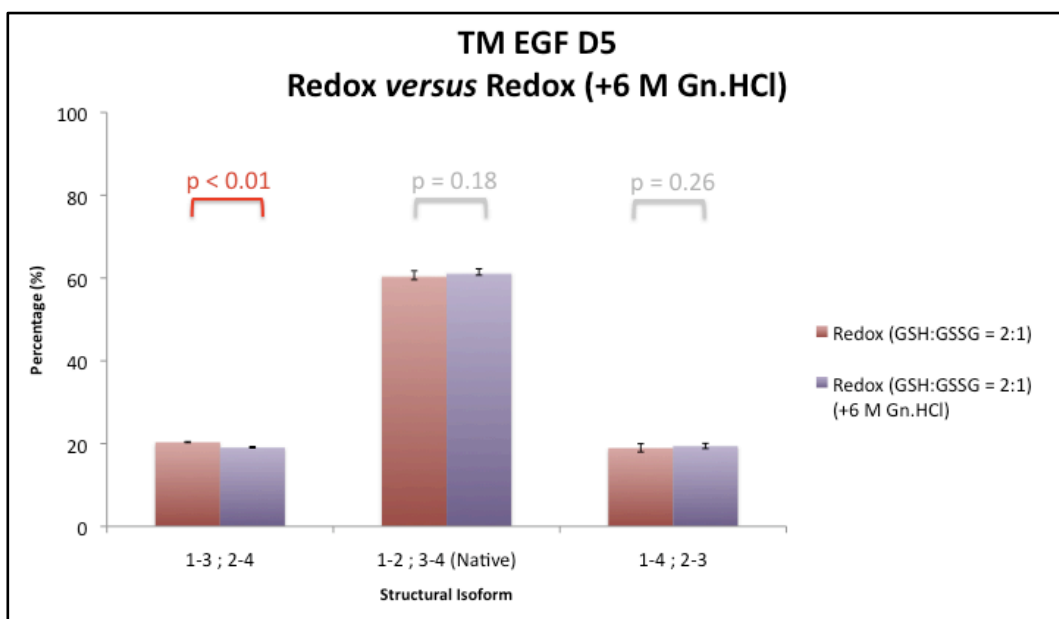


Figure 3.18 Pairwise comparison of t-TM EGF D5 structural isoform proportions obtained from redox reagent-mediated oxidation and redox reagent-mediated oxidation (with 6 M Gn.HCl) studies. Student's t-test (independent samples) using arcsine transformed-values were used for the calculation of probability (p)-values. Difference in proportion between corresponding structural isoforms is deemed to be significant when the p-value is less than 0.05 (one-tailed).

3.3.3.3 Putting the requirement for side-chain interactions into the context of nature's selection for t-TM EGF D4's and t-TM EGF D5's structural determinants

From the above discussion, it is clear that the EGF-like domain requires side-chain interactions to acquire its canonical fold with the C₁-C₃, C₂-C₄ disulfide-connectivity. As such, the structural determinants of t-TM EGF D4 was probably "chosen" to enable optimal participation in side-chain interactions so that the domain could preferentially fold into its C₁-C₃, C₂-C₄ isoform under normal oxidative conditions. When these structural determinants were prevented from interacting, the domain would not be able to fold into its native isoform and instead adopt an alternate isoform which does not require side-chain interactions to form.

On the other hand, since the folding of the C₁-C₂, C₃-C₄ isoform does not

require “information” from side-chain interactions, it is plausible to assume that t-TM EGF D5’s structural determinants, unlike that of t-TM EGF D4, are “selected” for optimal disengagement in side-chain interactions under normal oxidative conditions. Therefore, the preferential folding of t-TM EGF D5 into its native isoform would not be affected even when 6 M Gn.HCl was included in the oxidative folding buffer. In such a case, any side-chain interactions that occur in t-TM EGF D5 would not be the main determinant of its native fold and the associated disulfide-connectivity.

With regards to the C₁-C₄, C₂-C₃ isoform, the proportion of this particular isoform in t-TM EGF D4 had also been shown to increase significantly when oxidative folding was performed in the presence of denaturant. However, it did not manage to reach a proportion as high as that of the C₁-C₂, C₃-C₄ isoform. This is probably because of the shorter inter-cysteine loop between C₂ and C₃ which disfavors formation of a disulfide bond between these two cysteine residues due to steric hindrance/clashes. Thus, it remains to be seen whether a longer inter-cysteine loop would draw folding tendencies away from the C₁-C₃, C₂-C₄ and C₁-C₂, C₃-C₄ isoform. For this purpose, an EGF-like domain with a longer inter-cysteine loop between C₂ and C₃ would be needed for verification. An example of this would be EGF-like domain 2 of human thrombospondin-2 which has eight residues between C₂ and C₃ instead of just three in TM EGF D4 and TM EGF D5. Moreover, the short C₂-C₃ inter-cysteine loop of TM EGF D4 and D5 might be “nature’s strategy” to divert folding away from the C₁-C₄, C₂-C₃ isoform so that only binary decision between C₁-C₃, C₂-C₄ or C₁-C₂, C₃-C₄ is needed. This binary decision then

depends on whether their respective structural determinants participate in side-chain interactions or not.

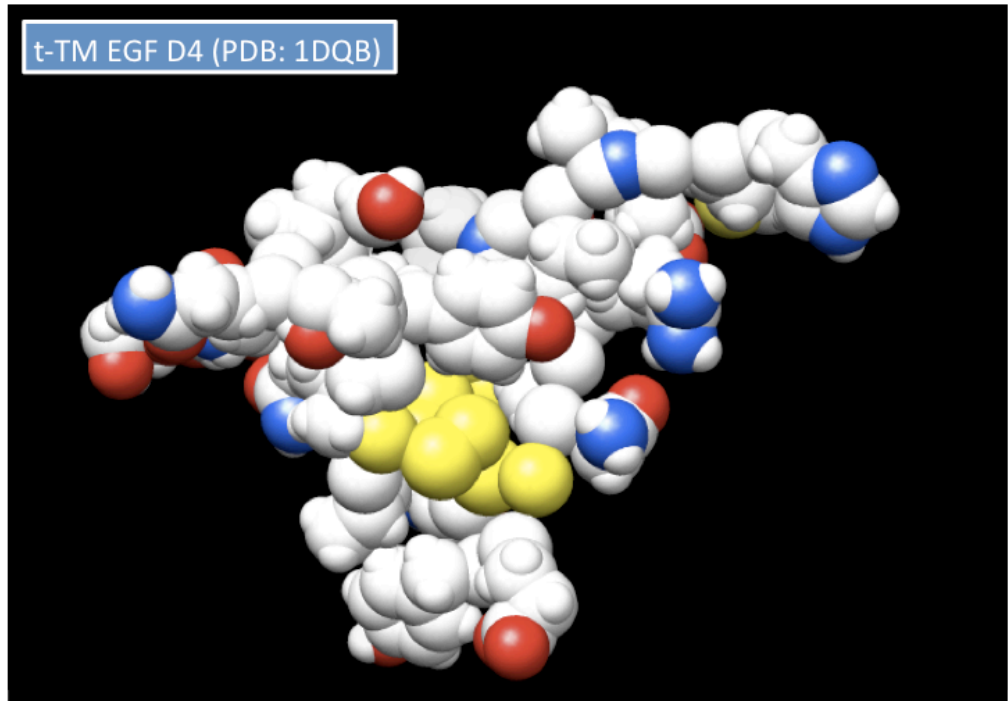
3.3.3.4 Probing into the nature of the side-chain interaction

Although side-chain interactions had been demonstrated to be necessary for the EGF-like domain to fold into its canonical C₁-C₃, C₂-C₄ fold, the nature of the side-chain interactions remained elusive. This is because the effects of Gn.HCl on side-chain interactions could not be differentiated between hydrophobic interaction or electrostatic interaction — i.e. The guanidinium ion could interact with hydrophobic side-chains to disrupt hydrophobic interactions [85, 86], and in addition, the ionic nature of Gn.HCl (i.e. guanidinium cation and chloride anion) could also mask any electrostatic interactions/repulsions present in the protein molecule [87].

Consideration for hydrophobic side-chain interactions as the force responsible for dictating the C₁-C₃, C₂-C₄ fold was based on the fact that compact protein structure is often stabilized by hydrophobic interactions [88]. Based on the solution structure of TM EGF D4-D5 solved by Wood, Sampoli Benitez and Komives [PDB: 1DQB] [79], it was observed that the t-TM EGF D4 segment folds into a rather compact structure (Figure 3.19A). Thus, hydrophobic interaction is likely to be involved in guiding the fold of the domain, as well as, dictating the C₁-C₃, C₂-C₄ disulfide-connectivity that reinforces the compact structure. On the other hand, the t-TM EGF D5 segment of the TM EGF D4-D5 structure folds into a less compact structure (Figure 3.19B) with most its side-chains facing away from the central core. This excludes hydrophobic

(A)

t-TM EGF D4 (PDB: 1DQB)



(B)

t-TM EGF D5 (PDB: 1DQB)

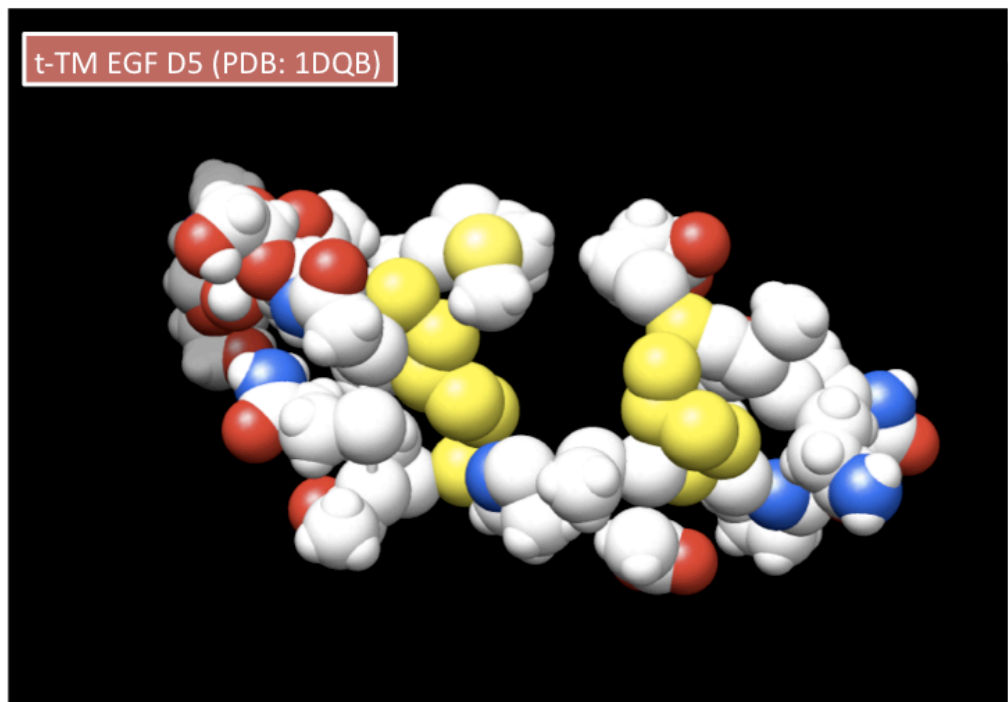


Figure 3.19 Space-filled model of (A) t-TM EGF D4 and (B) t-TM EGF D5. The model of these segments were extracted from PDB: 1DQB, which represents the solution structure of the TM EGF D4-D5 fragment solved by Wood, Sampoli Benitez and Komives (2000).

side-chain interaction as the dominant force in guiding the fold of the C₁-C₂, C₃-C₄ structural isoform.

The next consideration is that of electrostatic interactions. Although electrostatic interaction was also disrupted by high concentration of Gn.HCl, it was not likely the cause of t-TM EGF D4's failure to follow its native folding tendency in the presence of denaturant. This is because oppositely-charged residues in t-TM EGF D4 are all located in the N-terminal half of the truncated domain (Figure 3.1A), and are thus unlikely to be involved in guiding the formation of the overall compact structure under normal oxidative conditions.

As for the acidic t-TM EGF D5, it only contain three charged residues — i.e. one aspartic acid located on each side of C₃ and one glutamic acid N-terminal to C₄ (Figure 3.1B). These common charges might result in electrostatic repulsion that could bring C₃ and C₄ further away from each other than the equivalent cysteine residues in t-TM EGF D4. Indeed, a simple measurement of the C α -C α distance of C₃ and C₄ in t-TM EGF D4 and t-TM EGF D5 yield a distance of 4.828 Å and 5.400 Å, respectively. Structurally, disulfide bond formation between cysteine residues that are too close together (e.g. such as across two strands in a β -sheet [89]) creates strain and is thus unfavorable. Thus, the increased distance between C₃ and C₄ in t-TM EGF D5 might be more favorable for disulfide bond formation between these two residues to create the C₁-C₂, C₃-C₄ isoform. However, the 6 M Gn.HCl-experiments could disprove the above argument. The high content of guanidinium cation could possibly shield the negatively charged acidic residues in t-TM EGF D5 to reduce the effect of electrostatic repulsion. If electrostatic repulsion (and thus

lack of side-chain interactions) is responsible for the favorable formation of the C₃-C₄ disulfide bond, the proportion of the C₁-C₂, C₃-C₄ isoform would be brought down by the presence of Gn.HCl in the oxidative folding buffer. However, this did not happen and the C₁-C₂, C₃-C₄ remained the dominant isoform, with percentage values unaltered, in oxidative folding experiments conducted in the presence of 6 M Gn.HCl.

In view of these considerations, it seemed that hydrophobic side-chain interactions and lack thereof is responsible for the differential folding tendencies of t-TM EGF D4 and t-TM EGF D5, respectively. Thus, to confirm this conclusion, another set of experiments, using a different chemical reagent to manipulate the oxidative folding environment, was performed. Here, 0.5 M NaCl was chosen as the reagent — i.e. NaCl like Gn.HCl could disrupt electrostatic interactions [90], but unlike Gn.HCl which disrupts hydrophobic interactions, the presence of high NaCl concentration increases the hydrophobic effect in proteins (a phenomena that serves as a basis for hydrophobic interaction chromatography) [91]. Here, it was hypothesized that the inclusion of 0.5 M NaCl to the folding buffer would result in the increase of the C₁-C₃, C₂-C₄ structural isoform in both t-TM EGF D4 and t-TM EGF D5 as the presence of hydrophobic interactions favor the formation of this structural isoform.

3.4 Contribution of Hydrophobic Interactions in the Folding Tendencies of t-TM EGF D4 and t-TM EGF D5

To determine the role of hydrophobic interactions in dictating the folding tendency of t-TM EGF D4 and t-TM EGF D5, 0.5 M NaCl was included in the redox oxidative folding buffer to increase the hydrophobic effect, and to disrupt/mask any possible electrostatic interactions and repulsions. Alteration in the folding tendency of both domains was then noted to ascertain if hydrophobic interactions is the main contributor to the C₁-C₃, C₂-C₄ fold of the canonical EGF-like domain.

3.4.1 *In vitro* oxidative folding of t-TM EGF D4 in the presence of 0.5 M NaCl

Folding of t-TM EGF D4 in the presence of 0.5 M NaCl was performed using the redox buffer system and three monomeric isoforms were obtained (Figure 3.20A). Based on the relative yield of the respective structural isoforms (Table 3.3A: Orange columns), the result showed that t-TM EGF D4 still had a preference towards its native isoform when folded in the presence of high salt content.

3.4.2 *In vitro* oxidative folding of t-TM EGF D5 in the presence of 0.5 M NaCl

In the presence of 0.5 M NaCl, redox reagent-mediated folding of t-TM EGF D5 yielded three monomeric isoforms (Figure 3.20B). Like t-TM EGF D4, t-TM

EGF D5 still showed a folding preference towards its native isoform when folded in the presence of 0.5 M NaCl (Table 3.3B: Orange columns).

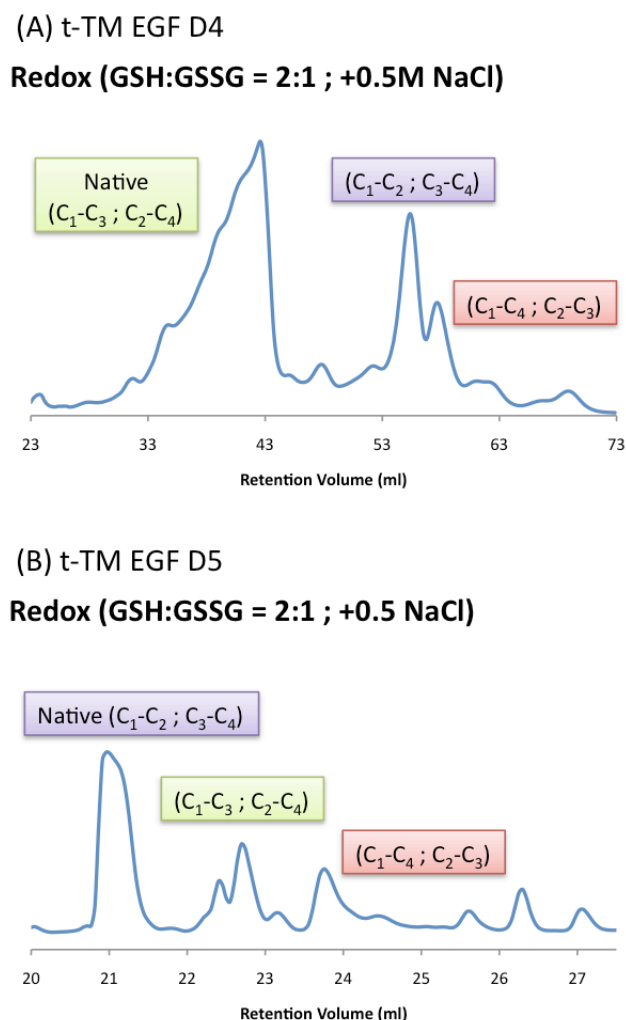


Figure 3.20 Analysis of (A) t-TM EGF D4 and (B) t-TM EGF D5 products obtained from redox reagent-mediated oxidation in the presence of 0.5 M NaCl by reversed-phase chromatography. Retention volume of the monomeric structural isoforms obtained were compared with that of regioselectively-synthesized structural isoforms for identification.

3.4.3 Hydrophobic interaction is necessary for the canonical C₁-C₃, C₂-C₄ fold of the EGF-like domain

As high salt content is known to disrupt electrostatic attraction, the preferential folding of t-TM EGF D4 into its native C₁-C₃, C₂-C₄ isoform even in the presence of 0.5 M NaCl showed that electrostatic interaction is not the nature

of the side-chain interaction involved in dictating the C₁-C₃, C₂-C₄ fold. Student's t-test comparison of corresponding t-TM EGF D4 structural isoforms obtained from oxidative folding in the absence *versus* presence of 0.5 M NaCl (Figure 3.21 and AP Table A3.9) showed a significant increase in the proportion of the native C₁-C₃, C₂-C₄ isoform in the experiments performed with NaCl. This was accompanied by significant decrease in the proportion of non-native structural isoforms (C₁-C₂, C₃-C₄ and C₁-C₄, C₂-C₃). Based on these results, increased in proportion of the C₁-C₃, C₂-C₄ isoform was deemed to be attributed to increased hydrophobic effect caused by the presence of 0.5 M NaCl. This reinforces the conclusion from Section 3.3.3.4 that the identity of the side-chain interactions involved in dictating the C₁-C₃, C₂-C₄ fold is that of hydrophobic interactions.

As for t-TM EGF D5, although it still folds predominantly into its native C₁-C₂, C₃-C₄ isoform in NaCl-containing folding buffer, pairwise comparison of corresponding structural isoforms obtained from oxidative folding in the absence *versus* presence of 0.5 M NaCl (Figure 3.22 and AP Table A3.10) showed a significant decrease in the proportion of its native C₁-C₂, C₃-C₄ isoform in experiments performed with NaCl. This decrease was accompanied by an increase in the canonical EGF-like C₁-C₃, C₂-C₄ isoform. This was probably due to salt-induced increase in hydrophobic effect which in turn make the folding into the C₁-C₃, C₂-C₄ conformer more favorable regardless of the exact identities of the EGF-like domains. In addition, as high salt content also mask electrostatic repulsion, the increased folding of t-TM EGF D5 into the compact C₁-C₃, C₂-C₄ isoform could also be attributed to this effect.

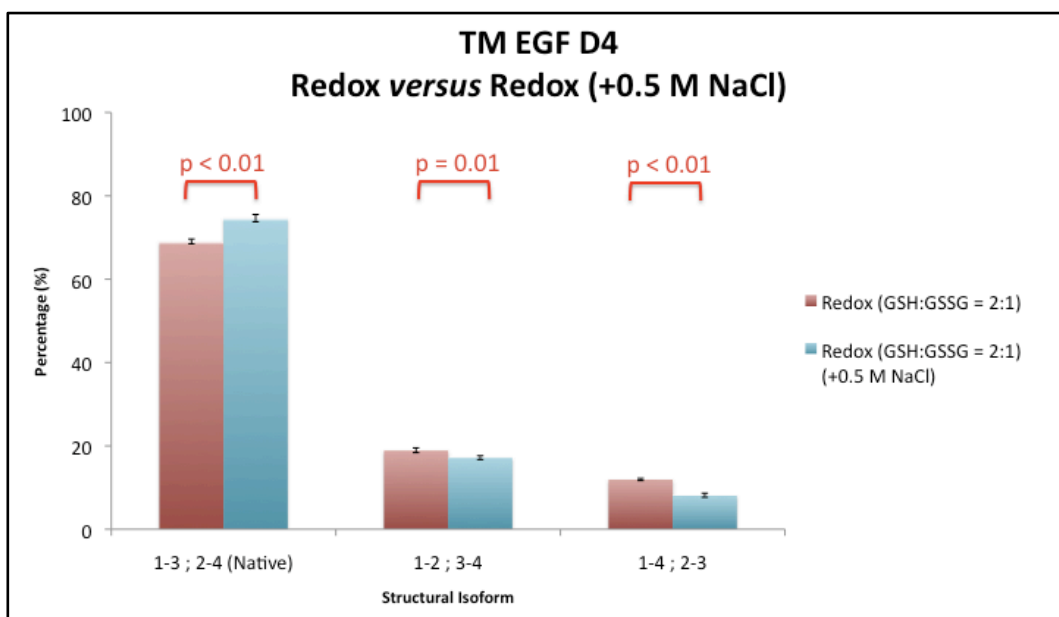


Figure 3.21 Pairwise comparison of t-TM EGF D4 structural isoform proportions obtained from redox reagent-mediated oxidation and redox reagent-mediated oxidation (with 0.5 M NaCl) studies. Student's t-test (independent samples) using arcsine transformed-values were used for the calculation of probability (p)-values. Difference in proportion between corresponding structural isoforms is deemed to be significant when the p-value is less than 0.05 (one-tailed).

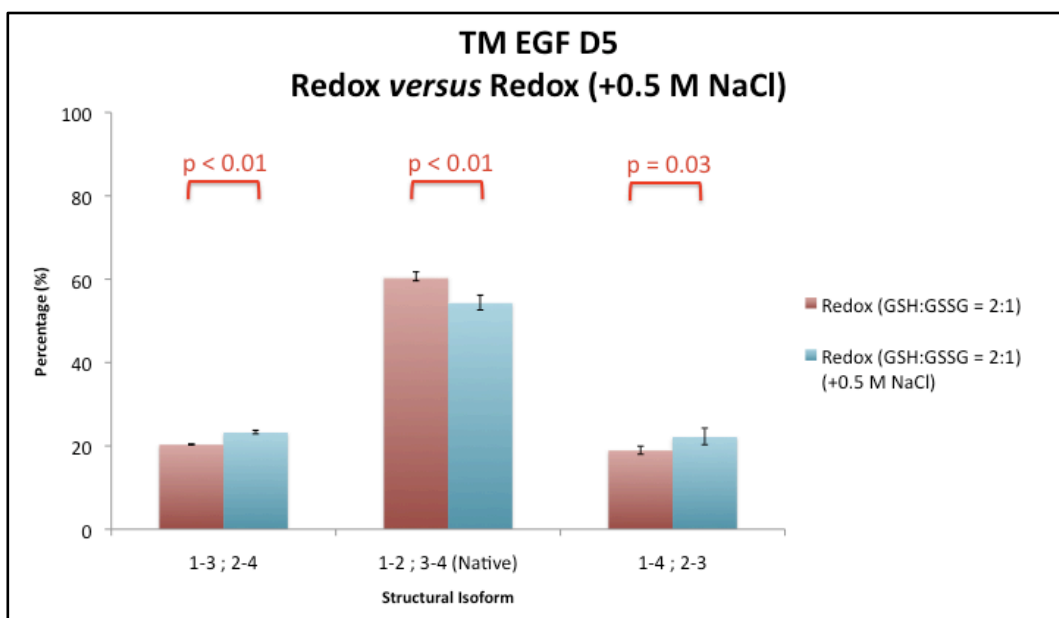


Figure 3.22 Pairwise comparison of t-TM EGF D5 structural isoform proportions obtained from redox reagent-mediated oxidation and redox reagent-mediated oxidation (with 0.5 M NaCl) studies. Student's t-test (independent samples) using arcsine transformed-values were used for the calculation of probability (p)-values. Difference in proportion between corresponding structural isoforms is deemed to be significant when the p-value is less than 0.05 (one-tailed).

However, in view of the observations made from the Gn.HCl-containing experiments, this is probably not the case.

To conclude, hydrophobic interactions had been identified as the dominant force that drives the C₁-C₃, C₂-C₄ fold of the canonical EGF-like domains. Revisiting what had been mentioned in Section 3.3.3.3, this meant that the structural determinants of the canonical EGF-like domain is “designed” to optimally engage specifically in hydrophobic interaction. On the contrary, the structural determinants of the C₁-C₂, C₃-C₄ isoform, are probably more polar (or less hydrophobic) and thus prefer to interact with the aqueous medium. Although polar side-chain interactions can also occur, the amino acid sequence of t-TM EGF D5 is probably “designed” such that balance of physical-chemical forces in the folding peptide could not outcompete the aqueous solvent for interaction with the side-chain of the structural determinants.

The identification of the dominant force that dictates the C₁-C₃, C₂-C₄ fold provided clues to the identity of the structural determinants in t-TM EGF D4. Thus, it is now of interest to identify the key hydrophobic residues in t-TM EGF D4 that is involved in dictating its canonical EGF-like fold.

3.5 Identification of Key Hydrophobic Residues as Structural Determinants of the Canonical EGF-like Domain fold in t-TM EGF D4

Based on the experimental evidence provided in Section 3.4, it can be suggested that the increase in hydrophobic interactions/effect generally increase folding into the C₁-C₃, C₂-C₄ isoform regardless of the exact identities of the EGF-like domains involved. In view of this generalized effect, the finding that hydrophobic interaction is the nature of the side-chain interactions that guides folding towards the C₁-C₃, C₂-C₄ fold could be applicable to other canonical EGF-like domains as well.

Therefore, for amino acids to satisfy the role of structural determinants in the canonical EGF-like domain, they have to be hydrophobic in nature. However, an addition requirement is that the amino acid residues at their equivalent positions in the non-canonical fold have to be either hydrophilic or less hydrophobic. This additional requirement assumes that the structural determinants for the C₁-C₂, C₃-C₄ fold are located in the same position along the amino acid sequence as that of the C₁-C₃, C₂-C₄ fold. This assumption is based on the following reasoning:

The EGF-like domain is an evolutionarily conserved modular unit with diverse functionality. Therefore, the positions of its structural determinants have to be conserved, while accommodating varied “functional” residues between them, to maintain the overall canonical fold across the domain family. Thus, the switch from canonical to non-canonical EGF-like domain fold in TM EGF D5 is more likely to be caused by a switch in chemical properties of the structural

determinants which are located at conserved position, rather than the “relocation” of structural determinants to cause a different fold. In the case of TM EGF D5, the C₁-C₂, C₃-C₄ fold is determined to be the result of the absence of side-chain interactions between key structural determinants.

To identify potential hydrophobic residues in t-TM EGF D4 for further studies, sequence alignment of t-TM EGF D4 and other canonical EGF-like domains from various proteins was performed to identify conserved hydrophobic residues. These other canonical EGF-like domains were chosen on the basis that their three-dimensional structures had been solved. Thus, EGF-like domains whose three-dimensional structures are unknown, but are assumed to possess the C₁-C₃, C₂-C₄ disulfide connectivity based on sequence homology, were not chosen.

Interestingly, results from the sequence alignment showed only one conserved hydrophobic/aromatic residue which is located two residues N-terminal to the C₄ residue (Figure 3.23). This hydrophobic/aromatic residue is also present in TM EGF D4 of other organisms, but is absent from TM EGF D5 (Figure 3.24). The amino acid in the equivalent position in TM EGF D5 is substituted by less hydrophobic residues. Although it seemed unlikely that a single residue is all that is needed to guide the folding of the EGF-like domain towards the C₁-C₃, C₂-C₄ conformer, this possibility could not be ruled out — Research on the structural determinants of α -conotoxin Iml showed that a mere switch from amide to acid at its C-terminal is enough to switch its disulfide-connectivity preference from C₁-C₃, C₂-C₄ to C₁-C₄, C₂-C₃ [41].

Here, the identified hydrophobic/aromatic residue satisfies the two criteria for

Sequence Alignment (ClustalW2) – EGF-like Domains of Various Proteins

```

P05979|32-70      -PVNPC---CYYPQOHQG---ICVRF-GLDR---YQCD 27
Q05769|18-55      --ANPC---CSNPCNRG---ECMST-GFDQ---YKCD 26
P35442|549-589    -PVDGC---LSNPCFPGA---QCSSF-PDGS---WSCG 27
P07204|365-405    -PVDPC---FRANC--EY---QQOPL-NQTS---YLCV 25
P08709|106-142    -DGDQC---ASSPCQNGG---SCKD--QLQS---YICF 26
P00742|86-122     -DGDQC---ETSPCQNGG---KCKD--GLGE---YFCT 26
P01133|972-1013   -SDSECP LSHDGYCLHDG---VCMYIEALDK---YACN 31
P01132|978-1019   -SYPGCPSSYDGYCLNGG---VCMHIESLDS---YTCN 31
Q6ULR6|44-83      --TASC---QDMSCSKQG---ECL--ETIGN---YTCG 25
P35442|590-647    -DLDECAL-VPDICEFSTSKVPRCVN--TQPG---FHCL 31
P46531|412-450    -DVDECSL-GANPCEHAG---KCTN--TLGS---FECQ 28
P46531|452-488    -DVNEC---VSNPCQND--TCLD--QIGE---PQCI 26
P46531|490-526    -NTDEC---ASSPOLHNG---RCLD--KINE---PQCE 26
P10493|384-424    -SQQTCAN-NRHQCSVHA---ECRD--YATG---PCCR 28
Q02297|177-239    SHLVKCAEKEKTFVNGG---ECFMVKDLNSPSRYLCK 35
P35442|648-692    -PENPCKD-KTHNCHKHA---ECLYLGHFSDP-MYKCE 32
      *          *          *          *          *

```

UniProt ID Residue number	Protein Name	Region	Organism
P05979 32-70	Prostaglandin G/H synthase 1	EGF-like	Ovis aries (Sheep)
Q05769 18-55	Prostaglandin G/H synthase 2	EGF-like	Mus musculus (Mouse)
P35442 549-589	Thrombospondin-2	EGF-like 1	Homo sapiens (Human)
P07204 365-405	Thrombomodulin	EGF-like 4	Homo sapiens (Human)
P08709 106-142	Coagulation factor VII	EGF-like 1	Homo sapiens (Human)
P00742 86-122	Coagulation factor X	EGF-like 1	Homo sapiens (Human)
P01133 972-1013	Pro-epidermal growth factor	EGF-like 9	Homo sapiens (Human)
P01132 978-1019	Pro-epidermal growth factor	EGF-like 9	Mus musculus (Mouse)
Q6ULR6 44-83	P-selectin	EGF-like	Homo sapiens (Human)
P35442 590-647	Thrombospondin-2	EGF-like 2	Homo sapiens (Human)
P46531 412-450	Neurogenic locus notch homolog protein 1	EGF-like 11	Homo sapiens (Human)
P46531 452-488	Neurogenic locus notch homolog protein 1	EGF-like 12	Homo sapiens (Human)
P46531 490-526	Neurogenic locus notch homolog protein 1	EGF-like 13	Homo sapiens (Human)
P10493 384-424	Nidogen-1	EGF-like 1	Mus musculus (Mouse)
Q02297 177-239	Pro-neuregulin-1, membrane-bound isoform	EGF-like	Homo sapiens (Human)
P35442 648-692	Thrombospondin-2	EGF-like 3	Homo sapiens (Human)

Figure 3.23 Sequence alignment of canonical EGF-like domains from various proteins. Shown here are the sequences of the EGF-like domain segment encompassing C₁ to C₄. The conserved hydrophobic/aromatic residue is highlighted in green. Conserved cysteine residues of the EGF-like domain are highlighted in yellow.

Sequence Alignment (ClustalW2) – TM EGF D4 and D5 from Various Organisms

t-TM EGF D4

```

P07204      PVDPCFRANCEYQCQPLNQTSYLCV
P06579      PVDPCFDNCEYQCQPVRSEHKCI
B3STX8      PVDPCFGTDCYEYEQVVGRTGYRCV
P15306      LLDPCFGSNCEFQCPVSPDYRCI
Q5W7P8      PVDPCFGSKCEYQCQPVSQTDYRCI
O35370      QLDPCFRSKCEYQCQPVNSTHYNCI
Q8HZ48      PLDPCFGTNCYQCLPLG-QNYRCI
      : * * * *      . * * : *      : .      : * :

```

t-TM EGF D5

```

P07204      MF--CNQTACPADC DPNTQ--ASCE
P06579      MF--CNQTS CPADCDPHYP--TICR
B3STX8      MF--CNQTS CPADCDPNKQ--DSCQ
P15306      MF--CNETSCPADCDPNSP--TVCE
Q5W7P8      MF--CNQTACPADC DPNSP--TSCQ
O35370      MF--CNETSCPADCDPNSP--SFCQ
Q8HZ48      MF--CNQTTCPADCDPNYP--STCL
      **      * * : * : * * * * * :      *

```

UniProt ID	Organism
P07204	Homo Sapiens (Human)
P06579	Bos taurus (Cattle)
B3STX8	Sus scrofa (Pig)
P15306	Mus musculus (Mouse)
Q5W7P8	Canis lupus familiaris (Dog)
O35370	Rattus norvegicus (Norway Rat)
Q8HZ48	Oryctolagus cuniculus (Rabbit)

Figure 3.24 Sequence alignment of t-TM EGF D4 and t-TM EGF D5 from various organisms. The conserved hydrophobic/aromatic residue in t-TM EGF D4 is highlighted in green. The less hydrophobic residues at the equivalent position in t-TM EGF D5 is highlighted in pink. The conserved cysteine residues of the EGF-like domain are highlighted in yellow.

being a structural determinant in the canonical fold of the EGF-like domain. To further verify its suitability as a structural determinant, the structures of these canonical EGF-like domains were inspected to see if this conserved residue is in hydrophobic contact with other residues in the domain. The analysis revealed that this conserved residue mainly makes hydrophobic contact with amino acid residues located within the first inter-cysteine loop (Figure 3.25). Examples of this contact within the canonical EGF-like domain is depicted in Figure 3.26 with EGF-like domain 1 of human coagulation factor VII (Figure 3.26A) and the EGF-like domain of human Pro-neuregulin-1 (Figure 3.26B).

When looking specifically at TM EGF D4 and D5, the conserved hydrophobic/aromatic residue of Tyr25 in TM EGF D4 is in close contact with Ala11 of the first inter-cysteine loop (Figure 3.27A). On the contrary, the amino acids residues at their equivalent positions (Figure 3.27B) in TM EGF D5 (Thr50 and Ala62) were not in contact (Figure 3.27C).

Although the contact between the conserved hydrophobic/aromatic residues and the identified residues in the first inter-cysteine loops might not represent hydrophobic contacts during the transition state of folding, these contacts are needed for the compact C₁-C₃, C₂-C₄ fold since they bring the third inter-cysteine loop (near the C-terminal) in close proximity to the first inter-cysteine loop (near the N-terminal). Thus, any disruption of these contacts would probably destabilize the C₁-C₃, C₂-C₄ structure to create the more loosely packed isoform with the C₁-C₂, C₃-C₄ disulfide-connectivity.

To experimentally verify the conserved hydrophobic/aromatic residue as the structural determinant in the canonical C₁-C₃, C₂-C₄ fold of the EGF-like

Identification of Interacting Residues – EGF-like Domains of Various Proteins

Inter-cysteine loop: 1 2 3

P05979 32-70	-PVN P C---CYY P CQH Q G---I V RF-GLDR---Y Q C D 27	
Q05769 18-55	--AN P C---CSN P CQNRG---E C MST-GFDQ---Y K C D 26	
P35442 549-589	-PVD G C---LSN P C F PGA---Q C SSF-PDGS---W S C G 27	
P07204 365-405	-PVD P C---FRAN C --EY---Q Q PL-NQTS---Y L C V 25	
P08709 106-142	-DGD Q C---ASS P CQNGG---S C KD--QLQS---Y I C F 26	
P00742 86-122	-DGD Q C---ETS P CQ Q G---K K D--GLGE---Y T C T 26	
P01133 972-1013	-SDSE C PLSHD G Y C LHDG---V C MYIEALDK---Y A C N 31	
P01132 978-1019	-SY P G C PS S YD G Y C LNGG---V C MHIESLDS---Y T C N 31	
Q6ULR6 44-83	--TAS C ---Q D MS C SK Q G---E C L--ETIGN---Y T C S 25	
P35442 590-647	-DL D E C AL-VPD I C F STSKVPCVN--T Q PG---F H C L 31	
P46531 412-450	-DVDE C SL- G AN P C E HAG---K C IN--TLGS---F E C Q 28	
P46531 452-488	-DVNE C ---VSN P CQ N DA---T C LD--Q I GE---F Q C I 26	
P46531 490-526	-NTDE C ---ASS P C L HNG---R C LD--K I NE---F Q C E 26	
P10493 384-424	-S Q QT C AN- N R H Q SVHA---E C RD--Y A TG---F C C R 28	
Q02297 177-239	SHLV K CAEKE K T F C V NGG---E C FMVKDLSNPS R Y L C K 35	
P35442 648-692	-PEN P C K D-K T H N C H KHA---E C Y L GHFSDP- M Y K C E 32	
	* * * *	

UniProt ID Residue number	Protein Name	Region	Organism
P05979 32-70	Prostaglandin G/H synthase 1	EGF-like	Ovis aries (Sheep)
Q05769 18-55	Prostaglandin G/H synthase 2	EGF-like	Mus musculus (Mouse)
P35442 549-589	Thrombospondin-2	EGF-like 1	Homo sapiens (Human)
P07204 365-405	Thrombomodulin	EGF-like 4	Homo sapiens (Human)
P08709 106-142	Coagulation factor VII	EGF-like 1	Homo sapiens (Human)
P00742 86-122	Coagulation factor X	EGF-like 1	Homo sapiens (Human)
P01133 972-1013	Pro-epidermal growth factor	EGF-like 9	Homo sapiens (Human)
P01132 978-1019	Pro-epidermal growth factor	EGF-like 9	Mus musculus (Mouse)
Q6ULR6 44-83	P-selectin	EGF-like	Homo sapiens (Human)
P35442 590-647	Thrombospondin-2	EGF-like 2	Homo sapiens (Human)
P46531 412-450	Neurogenic locus notch homolog protein 1	EGF-like 11	Homo sapiens (Human)
P46531 452-488	Neurogenic locus notch homolog protein 1	EGF-like 12	Homo sapiens (Human)
P46531 490-526	Neurogenic locus notch homolog protein 1	EGF-like 13	Homo sapiens (Human)
P10493 384-424	Nidogen-1	EGF-like 1	Mus musculus (Mouse)
Q02297 177-239	Pro-neuregulin-1, membrane-bound isoform	EGF-like	Homo sapiens (Human)
P35442 648-692	Thrombospondin-2	EGF-like 3	Homo sapiens (Human)

Figure 3.25 Identification of residues that interacts with the conserved hydrophobic/aromatic residues in various canonical EGF-like domains. Shown here are the sequences of the EGF-like domain segment encompassing C₁ to C₄. By inspection of the three-dimensional structure of the various EGF-like domains, amino acid residues which are in contact with the conserved/hydrophobic residue are identified. Here, the interacting residues are indicated in bold font. The conserved hydrophobic/aromatic residue is highlighted in green. The conserved cysteine residues of the EGF-like domain are highlighted in yellow.

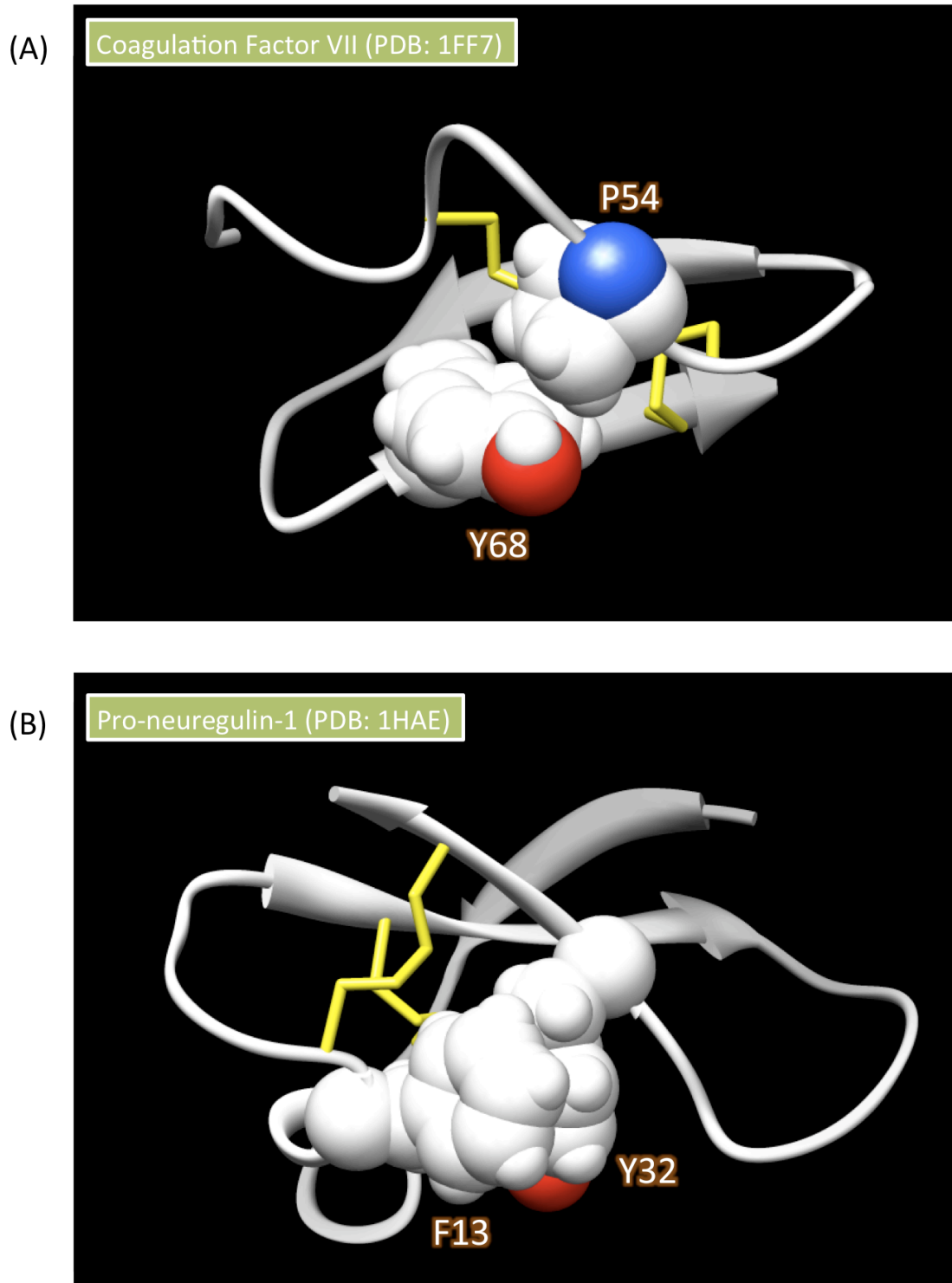


Figure 3.26 Residues interacting with the conserved hydrophobic/aromatic residues in (A) coagulation factor VII EGF-like domain 1 and (B) Pro-neuregulin-1 EGF-like domain. Depicted here are the EGF-like domain segments encompassing C₁ to C₄. The model of these segments were extracted from PDB: 1FF7 and PDB: 1HAE, respectively. Interacting residues were labeled, and the indicated positions are in accordance to the position numbers used in their respective PDB files.

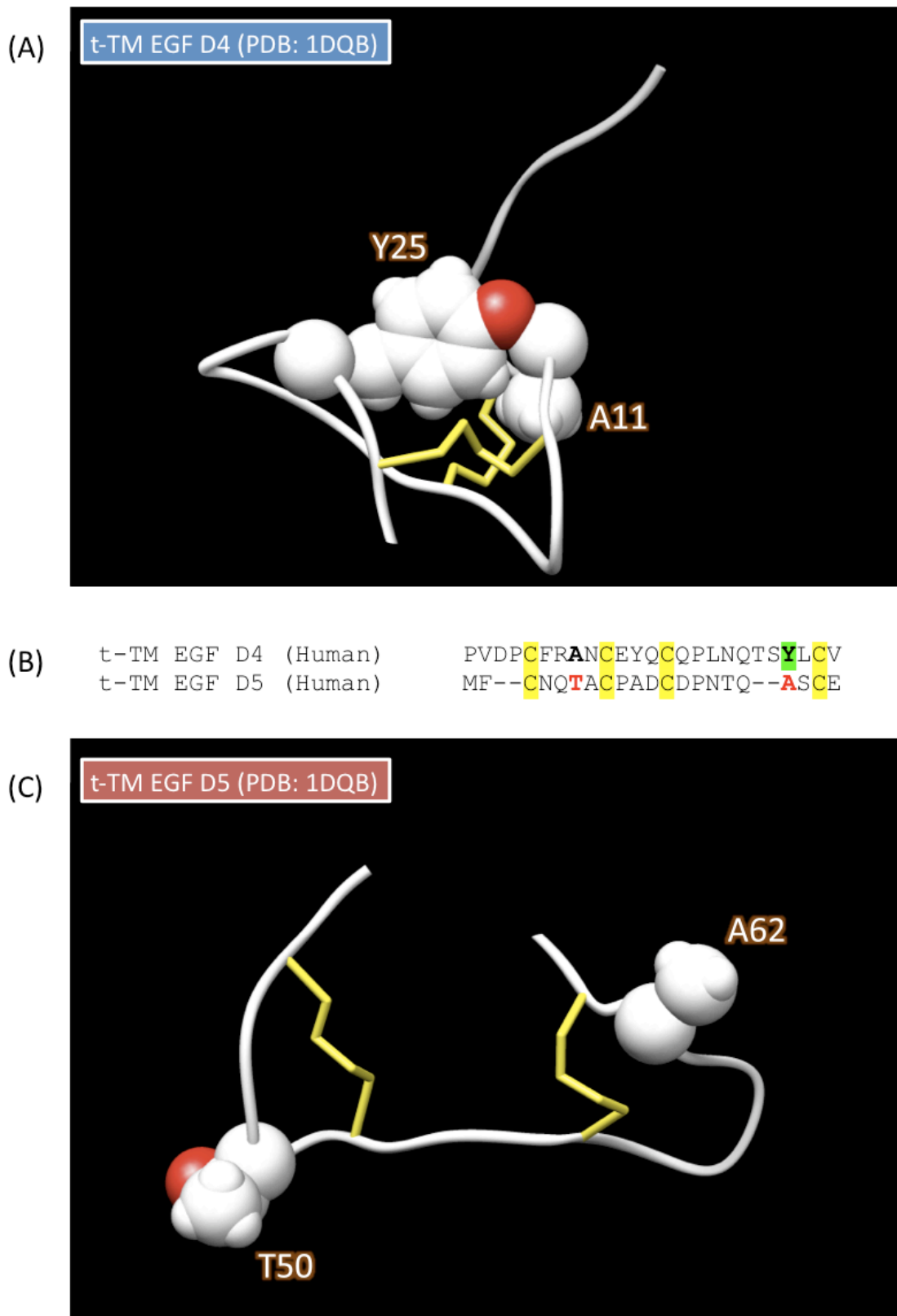


Figure 3.27 Residues interacting with the conserved hydrophobic/aromatic residue in the canonical EGF-like t-TM EGF D4. (A) Model of t-TM EGF D4 showing interaction between the conserved hydrophobic/aromatic residue of Y25 interacting with A11 of the first inter-cysteine loop. (B) Identification of the equivalent residues in t-TM EGF D5 which are indicated in (red) bold font. (C) Model of t-TM EGF D5 showing non-interaction between T50 and A62. The models of these segments were extracted from PDB: 1DQB. Positions of residues are labeled in accordance to the position numbers used in the PDB files.

domain, modified t-TM EGF D4 with Tyr25 substituted with threonine was synthesized. The threonine residue is more hydrophilic than the tyrosine residue as it lacks the hydrophobic aromatic ring of tyrosine. Moreover, it was chosen as a substitution for tyrosine as it also carries a hydroxyl group, thus making the hydrophilic substitution based solely on the removal of the hydrophobic aromatic ring.

The fully reduced modified t-TM EGF D4 peptide was then folded using air oxidation or redox reagent-mediated oxidation, either in the absence or presence of 6 M Gn.HCl. The dominant structural isoform obtained from these folding studies were then identified based on elution profile comparison with that of regioselectively-synthesized structural isoforms. If the conserved hydrophobic/aromatic residue is the structural determinant of the C₁-C₃, C₂-C₄ fold, the proportion of this conformer should decrease significantly in t-TM EGF D4 after its substitution with a more hydrophilic residue.

3.5.1 *In vitro* oxidative folding of t-TM EGF D4 (Y25T)

Oxidative folding of reduced t-TM EGF D4 (Y25T) was performed using air oxidation and redox reagent-mediated folding:

- (a) Air oxidation-mediated folding of reduced t-TM EGF D4 (Y25T) was completed in approximately 72 hrs as judged by the Ellman's test. Structural isoforms obtained from the reaction were resolved by reversed-phase chromatography and three monomeric isoforms were obtained (Figure 3.28).

(b) Folding of t-TM EGF D4 (Y25T) was performed in redox buffer over a period of 48 hrs. Like air oxidation, three monomeric isoforms were obtained (Figure 3.29).

In both oxidative folding studies, the retention volume of the three monomeric isoforms obtained matched well with that of regioselectively-synthesized structural isoforms. This enabled the identification of peaks in the chromatogram, and the relative proportions of the three isoforms were calculated (Table 3.4: Pink columns).

Results from both folding studies showed that t-TM EGF D4 has an altered folding preference after replacing the putative structural determinant of the canonical C₁-C₃, C₂-C₄ fold with a more hydrophilic residue — That is, instead of folding into the canonical fold of the EGF-like domain, t-TM EGF D4 (Y25T) displayed a folding preference towards the non-canonical C₁-C₂, C₃-C₄ isoform (Figure 3.30 and Figure 3.31).

Table 3.4 Percentages of structural isoforms obtained from oxidative folding of t-TM EGF D4 (Y25T) in various conditions

Oxidative folding of t-TM EGF D4 (Y25T)				
Structural Isoforms	(---) ^{a,b}		+6 M Gn.HCl ^c	
	Air Oxidation (%)	Redox Buffer System (%)	Air Oxidation (%)	Redox Buffer System (%)
C ₁ -C ₃ , C ₂ -C ₄ (Native)	22.78 ± 0.32	22.27 ± 0.42	8.16 ± 0.33	17.26 ± 0.45
C ₁ -C ₂ , C ₃ -C ₄	42.20 ± 0.36	42.25 ± 0.70	54.08 ± 0.52	45.92 ± 0.36
C ₁ -C ₄ , C ₂ -C ₃	35.02 ± 0.06	35.48 ± 0.78	37.76 ± 0.19	36.83 ± 0.18

^a Discussion in Section 3.5.1

^b Normal oxidative folding conditions (without denaturant or salt)

^c Discussion in Section 3.5.2

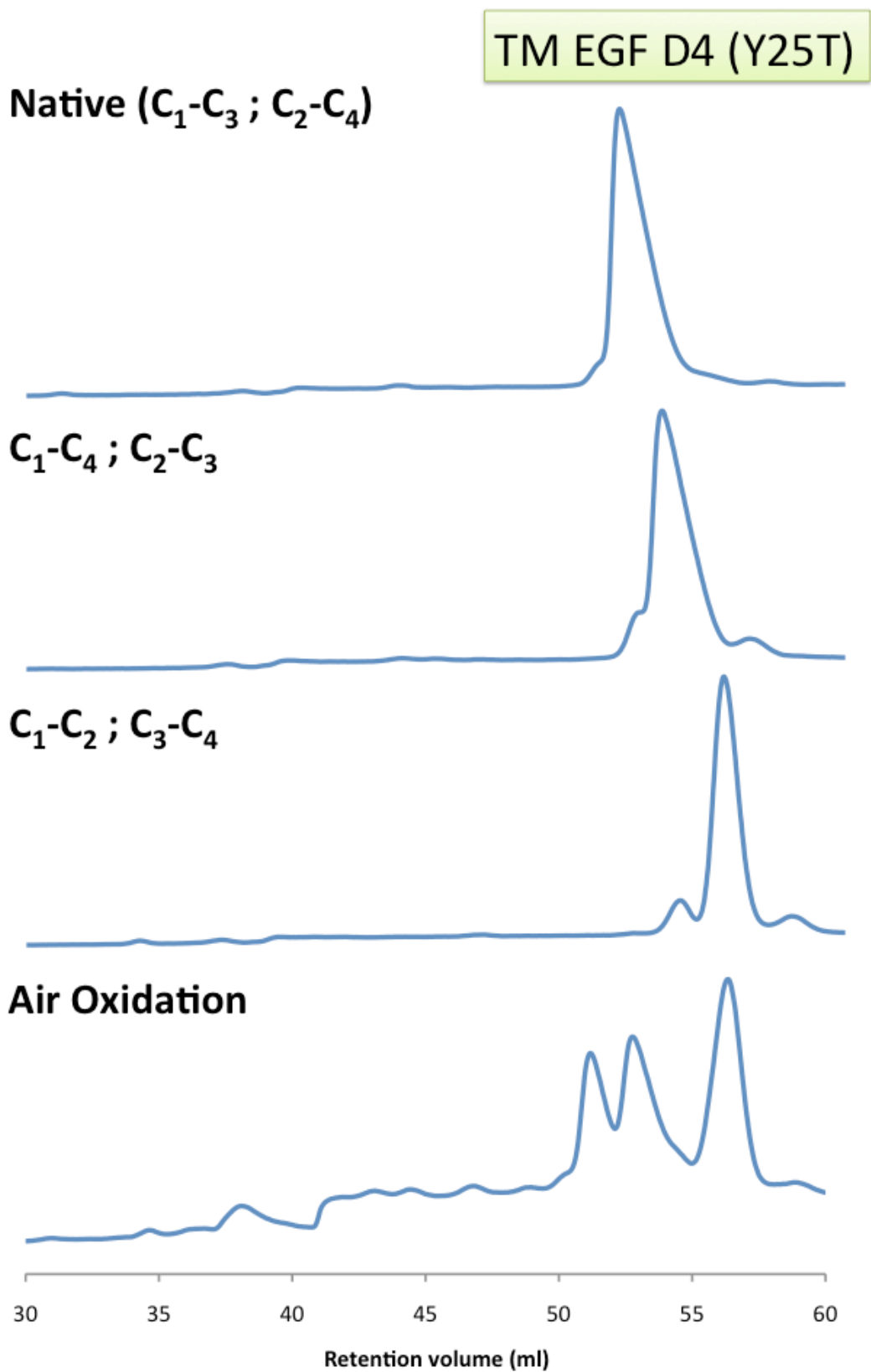


Figure 3.28 Analysis of t-TM EGF D4 (Y25T) air oxidation products by reversed-phase chromatography. Retention volume of the three monomeric structural isoforms obtained were compared with that of regioselectively-synthesized structural isoforms.

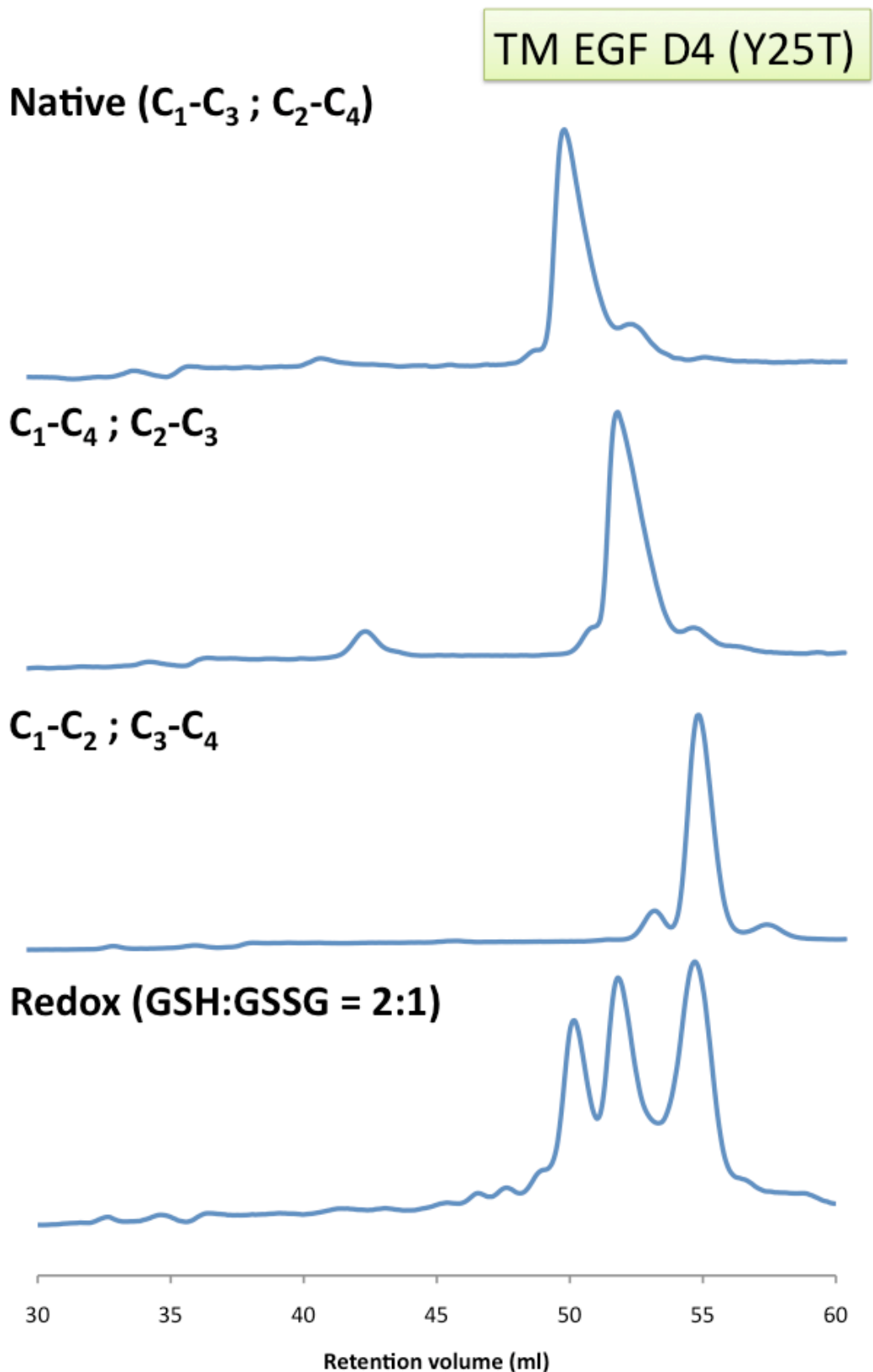


Figure 3.29 Analysis of t-TM EGF D4 (Y25T) redox reagent-mediated oxidation products by reversed-phase chromatography. Retention volume of the three monomeric structural isoforms obtained were compared with that of regioselectively-synthesized structural isoforms. Abbreviations: GSH — Reduced glutathione; GSSG — Oxidized glutathione.

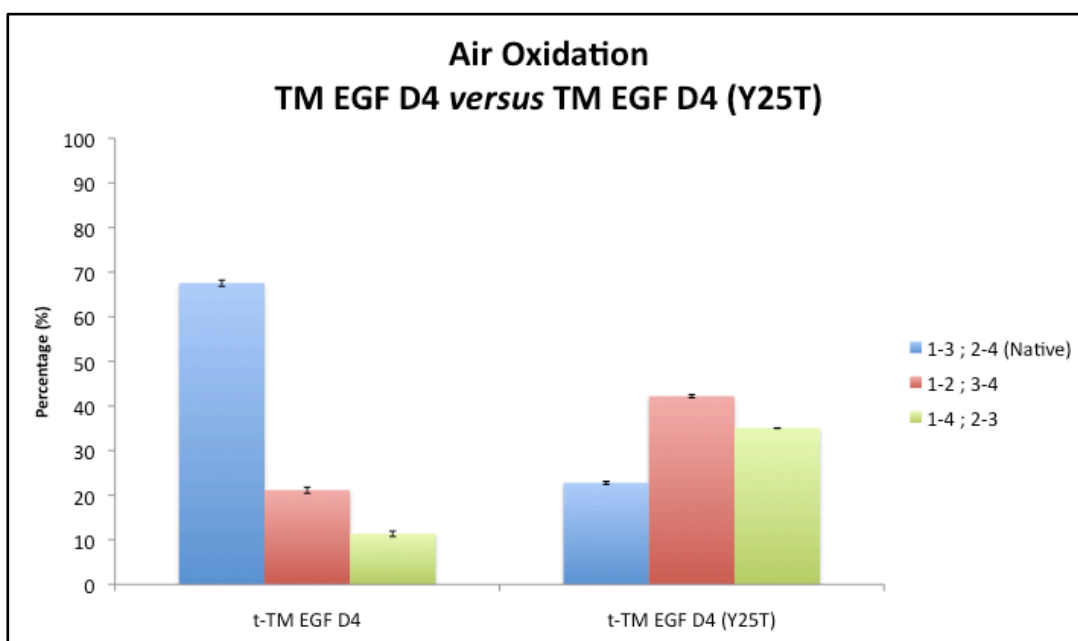


Figure 3.30 Proportion of structural isoforms obtained from air oxidation-mediated folding of t-TM EGF D4 and t-TM EGF D4 (Y25T). The dominant isoform obtained from t-TM EGF D4 is the canonical EGF-domain like (C₁-C₃, C₂-C₄) fold, while the dominant isoform obtained from t-TM EGF D4 (Y25T) is the non-canonical C₁-C₂, C₃-C₄ fold.

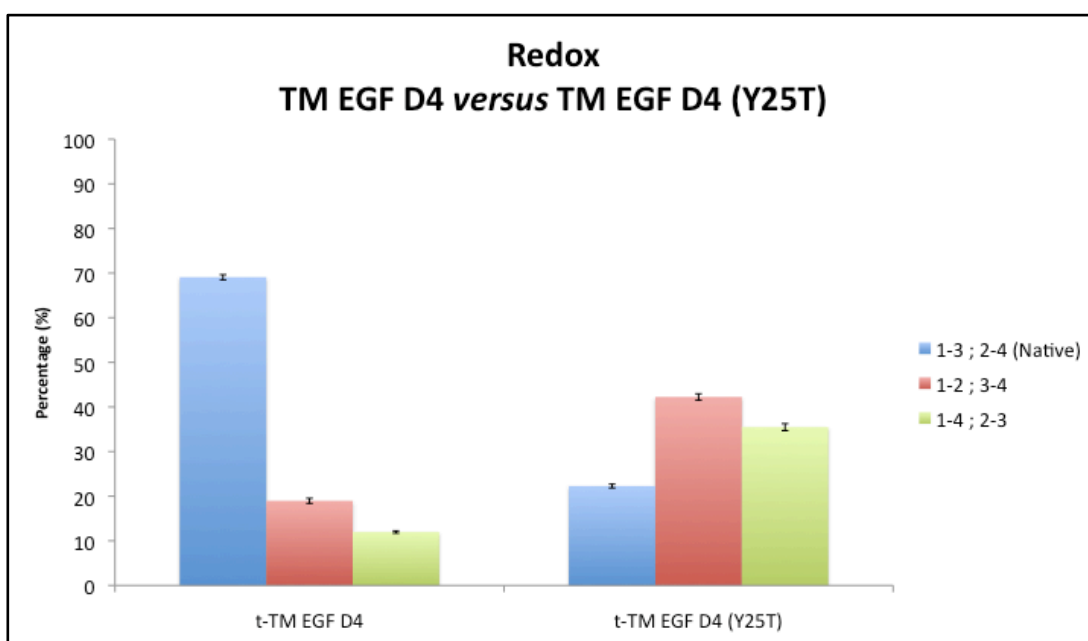


Figure 3.31 Proportion of structural isoforms obtained from redox reagent-mediated oxidative folding of t-TM EGF D4 and t-TM EGF D4 (Y25T). The dominant isoform obtained from t-TM EGF D4 is the canonical EGF-domain like (C₁-C₃, C₂-C₄) fold, while the dominant isoform obtained from t-TM EGF D4 (Y25T) is the non-canonical C₁-C₂, C₃-C₄ fold.

3.5.2 *In vitro* oxidative folding of t-TM EGF D4 (Y25T) in the presence of 6 M Gn.HCl

Air oxidation and redox reagent-mediated folding of t-TM EGF D4 (Y25T) was performed with the inclusion of 6 M Gn.HCl in the folding buffer:

- (a) For air oxidation in the presence of the denaturant, the reaction was completed in approximately 48 hrs as judged by the Ellman's test. Analysis by reversed-phase chromatography revealed that all three monomeric isoforms were obtained (Figure 3.32).
- (b) Redox reagent-mediated oxidative folding of t-TM EGF D4 (Y25T) in the presence of denaturant was allowed to proceed for 48 hours. It also yielded three monomeric isoforms (Figure 3.33).

The relative proportion of the three structural isoforms obtained in both studies revealed that the folding tendency of t-TM EGF D4 (Y25T) was unaltered despite the presence of 6 M Gn.HCl in the folding buffer (Table 3.4: Blue columns). Interestingly, this observation was similar to that of t-TM EGF D5, where the presence of denaturant in the folding buffer did not change the folding tendency of this non-canonical EGF-like domain.

3.5.3 The hydrophobic/aromatic residue, Tyr25, as the main structural determinant of t-TM EGF D4

The relative proportion of the three structural isoforms obtained from the oxidative folding of t-TM EGF D4 in the presence of 6 M Gn.HCl was similar to that of t-TM EGF D4 (Y25T) folded under normal oxidative conditions (Figure 3.34 and Figure 3.35). This suggests that hydrophobic interactions mediated

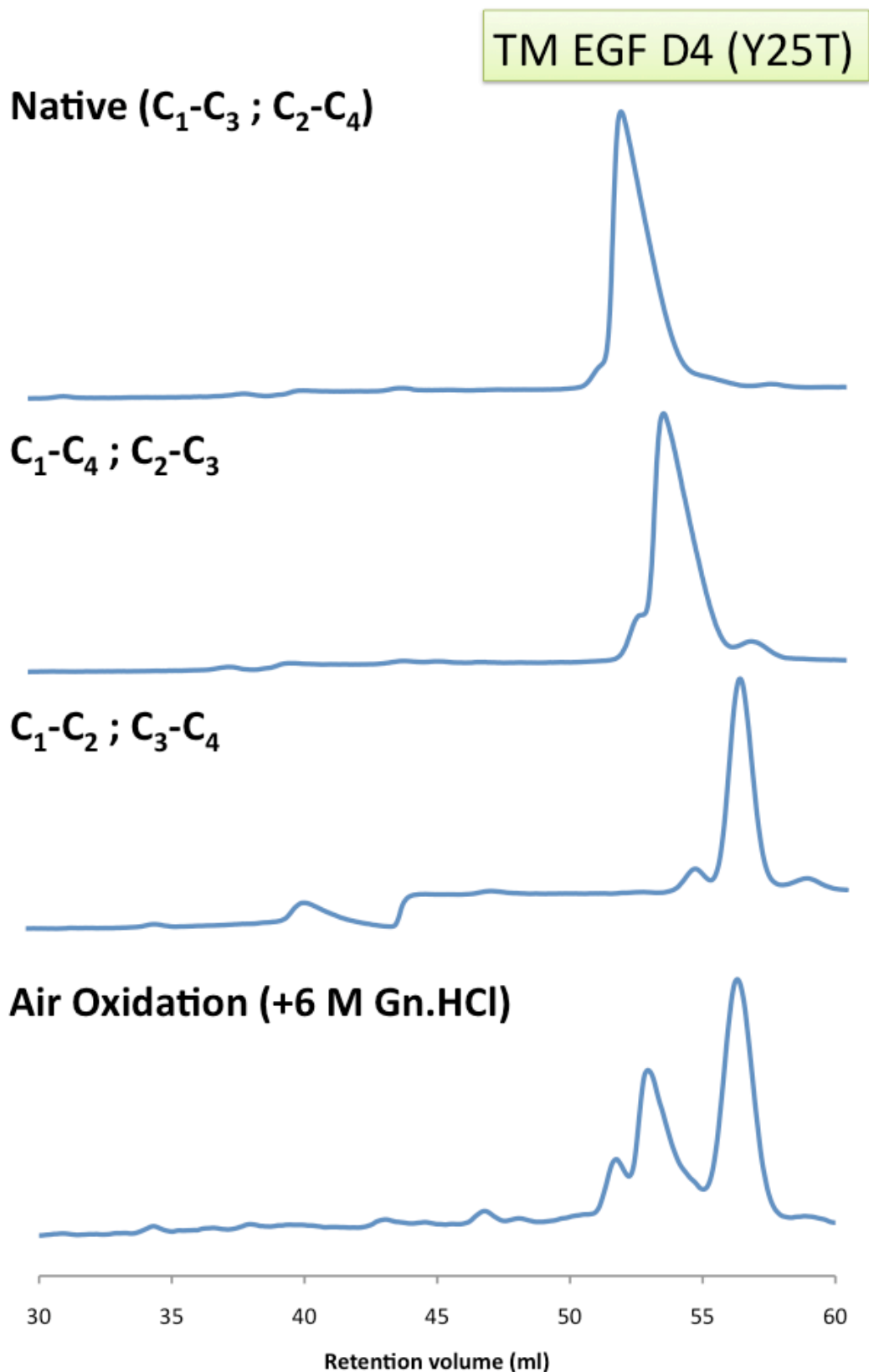


Figure 3.32 Analysis of t-TM EGF D4 (Y25T) products obtained from air oxidation in the presence of 6 M Gn.HCl by reversed-phase chromatography. Retention volume of the three monomeric structural isoforms obtained were compared with that of regioselectively-synthesized structural isoforms.

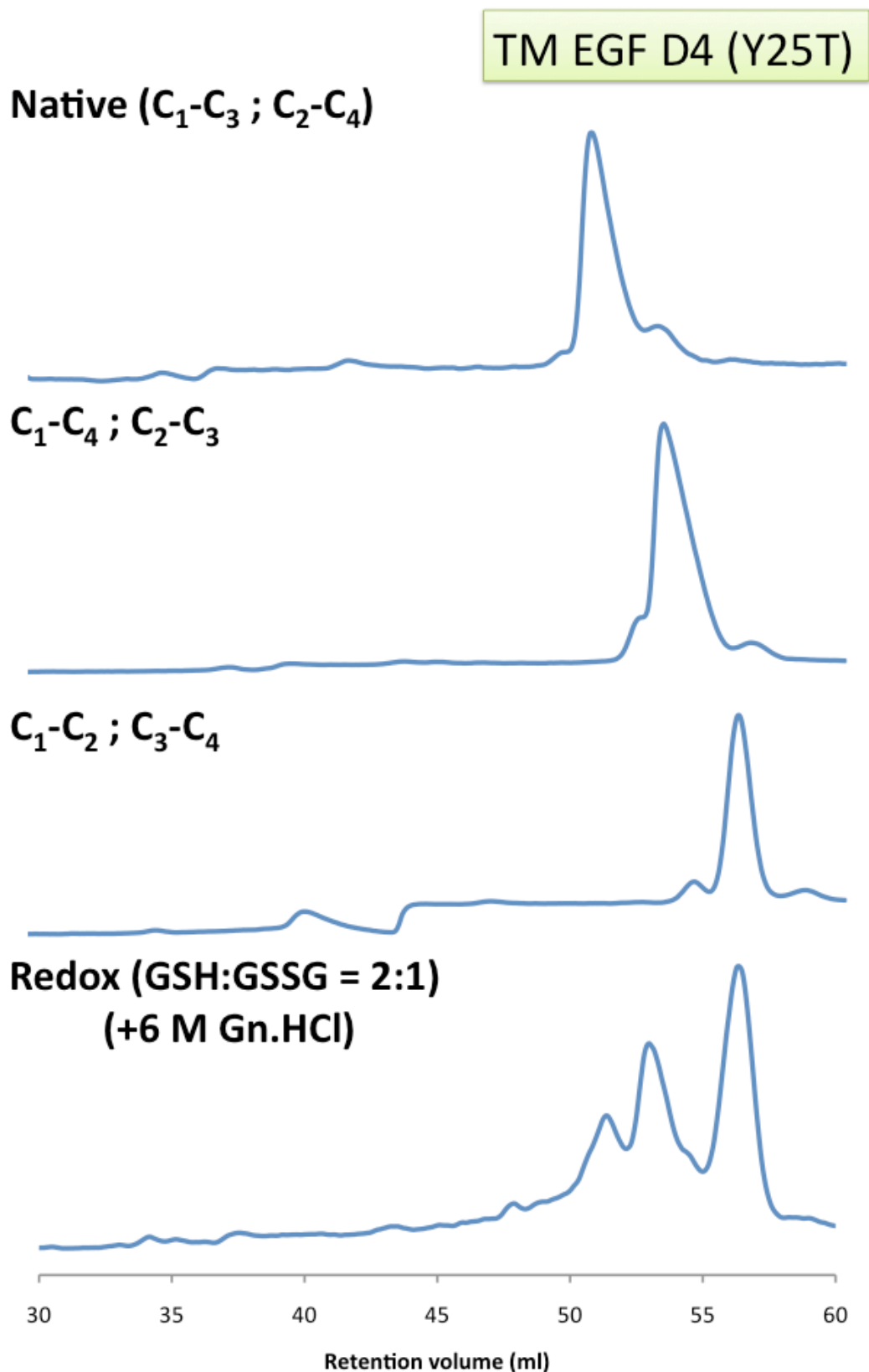


Figure 3.33 Analysis of t-TM EGF D4 (Y25T) products obtained from redox reagent-mediated oxidation in the presence of 6 M Gn.HCl by reversed-phase chromatography. Retention volume of the three monomeric structural isoforms obtained were compared with that of regioselectively-synthesized structural isoforms. Abbreviations: GSH — Reduced glutathione; GSSG — Oxidized glutathione.

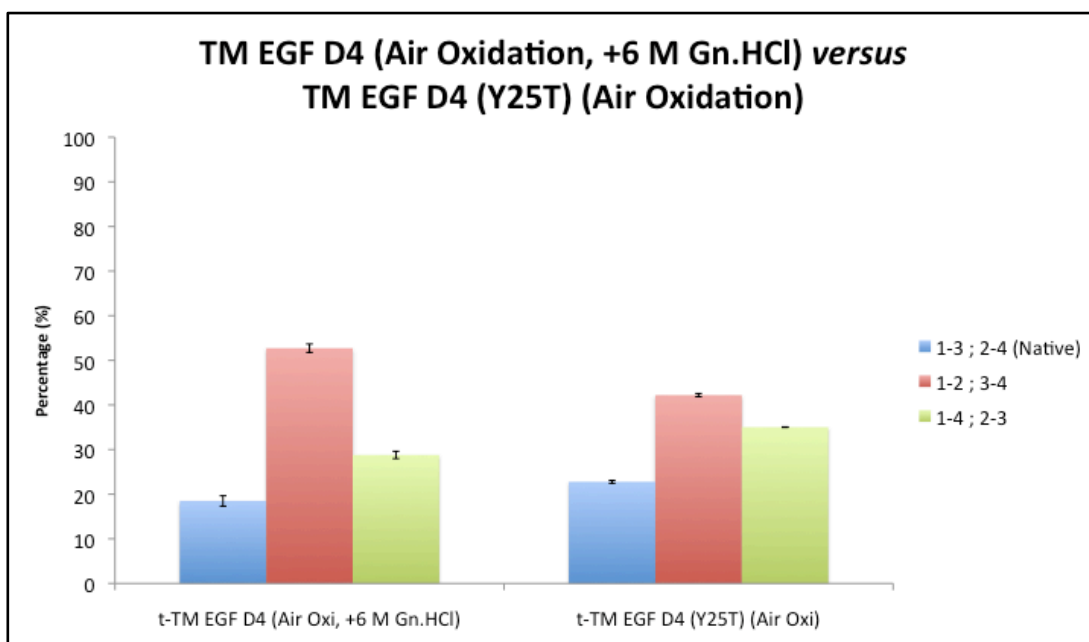


Figure 3.34 Proportion of structural isoforms obtained from air oxidation-mediated folding of t-TM EGF D4 (+6 M Gn.HCl) and t-TM EGF D4 (Y25T).

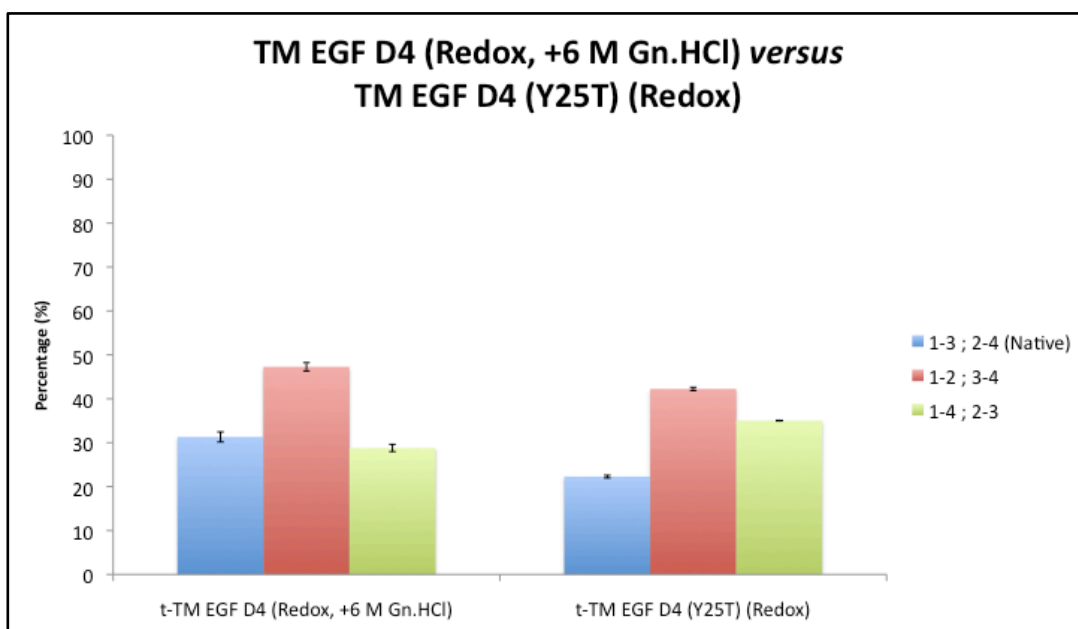


Figure 3.35 Proportion of structural isoforms obtained from redox reagent-mediated oxidative folding of t-TM EGF D4 (+6 M Gn.HCl) and t-TM EGF D4 (Y25T).

Tyr25 of t-TM EGF D4 was the identity of the side-chain interactions that were being disrupted by 6 M Gn.HCl, resulting in the shift of folding tendency from C₁-C₃, C₂-C₄ (canonical EGF-like) to C₁-C₂, C₃-C₄. Moreover, the proportion of the canonical C₁-C₃, C₂-C₄ isoform obtained from the folding of t-TM EGF D4 (Y25T) and t-TM EGF D5 under normal oxidative conditions, and t-TM EGF D4 in the presence of denaturant were similar (Figure 3.36 and Figure 3.37). This observation provide further evidence that the conserved hydrophobic/aromatic residue is the main structural determinant of the canonical C₁-C₃, C₂-C₄ EGF-like domain fold, with the disruption of which leading to an alternate fold that does not require hydrophobic interactions to form.

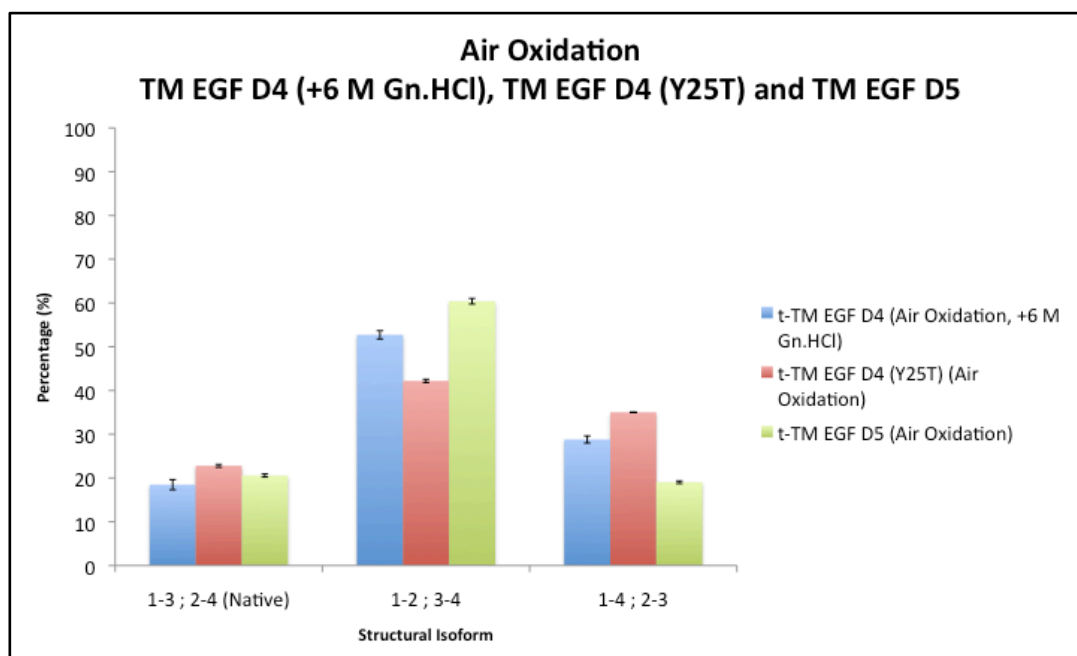


Figure 3.36 Comparison of structural isoform proportions obtained from air oxidation-mediated folding of t-TM EGF D4 (+6 M Gn.HCl), t-TM EGF D4 (Y25T) and t-TM EGF D5.

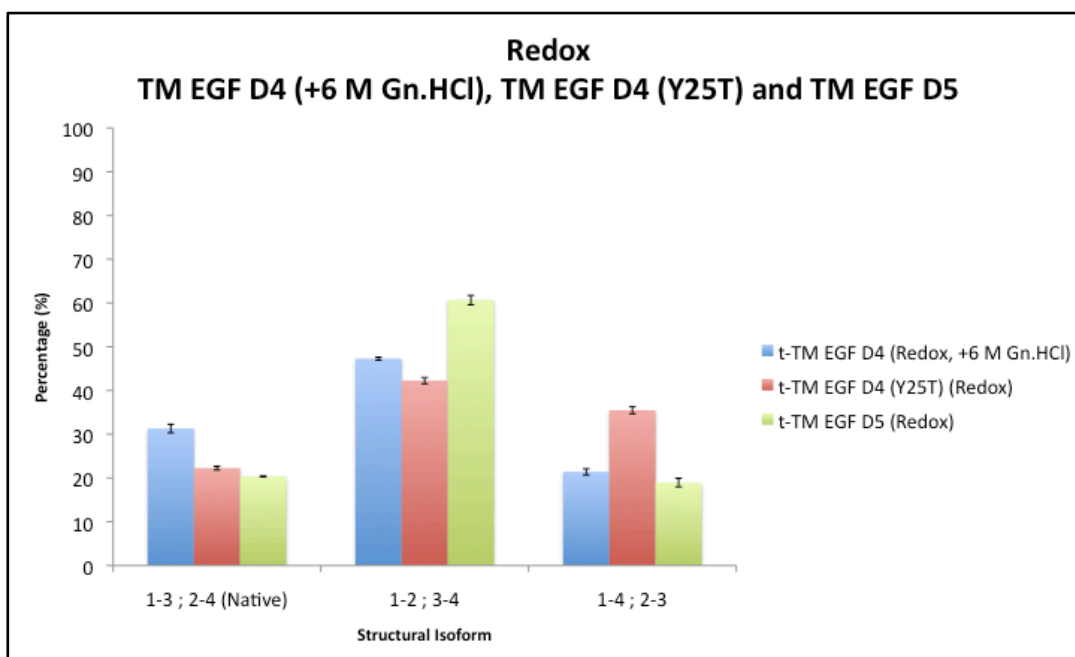


Figure 3.37 Comparison of structural isoform proportions obtained from redox reagent-mediated oxidative folding of t-TM EGF D4 (+6 M Gn.HCl), t-TM EGF D4 (Y25T) and t-TM EGF D5.

However, it should be noted that although the relative proportion of the non-canonical C₁-C₂, C₃-C₄ isoform increased when t-TM EGF D4 was folded in the presence of 6 M Gn.HCl or when t-TM EGF D4 (Y25T) was folded under normal oxidative conditions, its level did not reach as high as that of t-TM EGF D5 (Figure 3.36 and Figure 3.37). This suggests that t-TM EGF D5 contains its own specific structural determinants for the non-canonical C₁-C₂, C₃-C₄ fold in addition to the lack of the conserved hydrophobic/aromatic residue that is needed for the canonical EGF-like domain fold.



Chapter 4: Conclusion



4.1 Conclusion

The proposal of the thermodynamic hypothesis by Nobel Prize Laureate C.B. Anfinsen in the 1960s [92] had fueled intensive research with aims to decipher the protein folding code. However, despite such efforts, only fragmentary information, which consists mainly of general principles, had been obtained over the years. The presence of gaps in our current knowledge of the protein folding code had motivated the work described in this thesis. Here, the folding code of the evolutionarily conserved EGF-like domain is studied to provide more insights into how an amino acid sequence is being interpreted to result in three-dimensional structural information.

The EGF-like domain is a ubiquitous modular unit with diverse biological functions. All EGF-like domains consist of six conserved cysteine residues, with distinct hypervariability in the amino acid sequence of their inter-cysteine region. Although this hypervariability could explain the functional diversity of the various EGF-like domains, it has contributed to the puzzling question of how most EGF-like domains fold into their canonical C₁-C₃, C₂-C₄, C₅-C₆ scaffold despite of the inconsistency in sequence information.

A solution to this problem would involve the presence of conserved “structural determinants” embedded in the amino acid sequence of the EGF-like domain. These structural determinants could explain how the canonical three-looped structure of EGF-like domain is maintained in the midst of functional diversification. To find out the nature of these structural determinants, TM EGF D4 and TM EGF D5 were used as models for the study.

As TM EGF D4-D5 is the smallest co-factor active fragment of TM, interest in its structure-function relationship had resulted in interesting findings with regards to the structure of TM EGF D5. While TM EGF D4 folds into the canonical C₁-C₃, C₂-C₄, C₅-C₆ structure of the EGF-like domain, TM EGF D5 does not. Instead, it folds into an alternate conformation stabilized by the C₁-C₂, C₃-C₄, C₅-C₆ disulfide-connectivity. So how did this switch in conformation, from C₁-C₃, C₂-C₄ to C₁-C₂, C₃-C₄, occur? This switch could be attributed to a change in the physical-chemical properties of the canonical EGF-like domain's structural determinants which will be manifested as a difference in inter-molecular force, thus affecting the overall thermodynamic property of the polypeptide chain to result in a different fold. Based on this reasoning, the relative contribution of various inter-molecular forces in the folding of TM EGF D4 and TM EGF D5 was determined to provide clues to the identity of the structural determinants involved.

The first objective of this thesis was to narrow down the region where the structural determinants are located. Since the structural/disulfide-connectivity difference between TM EGF D4 and D5 lies in the first two disulfide bonds within their N-terminal segments (encompassing C₁ to C₄), it was of interest to see whether the folding information is located locally within that segment or non-locally in the C-terminal segment (encompassing C₅ to C₆). To this end, fully reduced, truncated versions of both domains were synthesized (t-TM EGF D4 and t-TM EGF D5) so that the oxidative folding of both domains could be performed without their respective C-terminal segments.

With the aid of regioselectively synthesized structural isoforms for peak identification in oxidative folding studies, results obtained from air oxidation and redox reagent-mediated oxidation studies showed that t-TM EGF D4 and t-TM EGF D5 still fold preferentially into their respective native structural isoforms despite the absence of the C-terminal segment. This suggests that the structural determinants of both domains lie locally within their N-terminal segments, encompassing C₁ to C₄.

The next objective was to determine the relative contribution of side-chain interactions to the folding tendency of both domains. Here, 6 M Gn.HCl was included in the oxidative folding experiments to disrupt any side-chain interactions present within the folding peptide. This changed the folding tendency of t-TM EGF D4 from that of its native C₁-C₃, C₂-C₄ conformer to that of the C₁-C₂, C₃-C₄ conformer. On the contrary, the disruption of side-chain interactions did not affect the folding tendencies of t-TM EGF D5 at all, and even resulted in a slight decrease of the C₁-C₃, C₂-C₄ isoform. These observations suggested that side-chain interactions are needed to guide the fold of EGF-like domains towards its canonical C₁-C₃, C₂-C₄ conformer. If side-chain interactions are absent, the default conformation adopted would be that of the C₁-C₂, C₃-C₄ conformer. When these findings were put into perspective, this meant that the structural determinants of the C₁-C₃, C₂-C₄ fold are selected for optimal engagement in side-chain interactions, while the converse is true for the C₁-C₂, C₃-C₄ fold.

Although Gn.HCl disrupts side-chain interactions, its effect could not be differentiated between the disruption of electrostatic interactions or

hydrophobic interactions. Thus, 0.5 M NaCl was included in the oxidative folding experiments to disrupt any electrostatic interactions, as well as, to increase the hydrophobic effect within the folding peptides. Unlike 6 M Gn.HCl, the inclusion of 0.5 M NaCl did not alter the folding tendency of t-TM EGF D4, and even increased the proportion of its native C₁-C₃, C₂-C₄ structural isoform. Therefore, this suggests that electrostatic interactions is not the nature of the side-chain interactions disrupted by 6 M Gn.HCl and that the increase in the canonical C₁-C₃, C₂-C₄ isoform was attributed to the increased hydrophobic effect. As for t-TM EGF D5, the presence of 0.5 M NaCl also did not change its folding tendency, but an increase in the proportion of the C₁-C₃, C₂-C₄ isoform was observed. These results collectively identifies the role of hydrophobic interactions in guiding the fold of the EGF-like domain towards the C₁-C₃, C₂-C₄ fold.

The final objective of this thesis was to identify key hydrophobic residues as the structural determinants of the canonical EGF-like domain fold. A sequence alignment of canonical EGF-like domains from various proteins helped identify a conserved hydrophobic/aromatic residue which is located two residues N-terminal to the C₄ residue. Interestingly, this conserved hydrophobic/aromatic residue is not present in its corresponding position in t-TM EGF D5. When the structures of various canonical EGF-like domains were examined, this conserved hydrophobic/aromatic residue mainly make contacts with residues in the first inter-cysteine loop of the domain. In TM EGF D4, this contact is present between Ala11 and Tyr25. However, this contact is not present in the equivalent positions in t-TM EGF D5 (Thr6 and Ala18).

With these analysis, an attempt was made to verify this hydrophobic/aromatic residue as the structural determinant of the canonical fold of the EGF-like domain. To this end, the Tyr25 residue of t-TM EGF D4 was substituted with a more hydrophilic threonine residue. If the conserved hydrophobic/aromatic residue is a structural determinant of the C₁-C₃, C₂-C₄ fold, the placement of this residue into t-TM EGF D4 should decrease its folding towards the C₁-C₃, C₂-C₄ conformer. Indeed, when t-TM EGF D4 (Y25T) was folded under oxidative conditions, with and without denaturant, it displayed a preferential folding towards the non-canonical C₁-C₂, C₃-C₄ conformer. More importantly, this was accompanied by a sharp drop in the proportion of the canonical C₁-C₃, C₂-C₄ conformer. This suggests that the conserved hydrophobic/aromatic residue is indeed the main structural determinant of the canonical C₁-C₃, C₂-C₄ fold of the EGF-like domain.

4.2 Future Work

4.2.1 Verifying the structural determinant of the canonical EGF-like domain fold

Future work for the current study would involve the continued focus on the verification of the conserved hydrophobic/aromatic residue as the structural determinant of canonical EGF-like domain fold. To this end, an alternate strategy involving the insertion of the hydrophobic/aromatic residue into t-TM EGF D5 at its equivalent position is proposed. In support of the current evidences, this insertion is expected to increase the proportion of the canonical C₁-C₃, C₂-C₄ fold in t-TM EGF D5.

4.2.2 The role of the structural determinant in the transition state of protein folding

After confirming the identity of the structural determinant, its role in the transition state of protein folding should be examined. For t-TM EGF D4, the slow kinetics of oxidative folding and the unique chemistry of the disulfide bond meant that folding intermediates could be trapped in a time course manner using either chemical modification (of free thiol groups) or acid-trapping. The structures of these trapped intermediates could then be analyzed by NMR to examine the contacts (i.e. native or non-native) made by this structural determinant during the process of folding.

4.2.3 Extending the study to other canonical EGF-like domains

When t-TM EGF D4 and t-TM EGF D5 were folded in the presence of 6 M Gn.HCl, the loss of side-chain interactions disfavored folding into the C₁-C₃, C₂-C₄ isoform for both domains. On the contrary, the increased hydrophobic effect mediated by 0.5 M NaCl drove up the proportion of C₁-C₃, C₂-C₄ in both domains. These generalized effect, without regards to the exact identities of the EGF-like domain, meant that the conclusion regarding hydrophobic interactions as the dominant driving force in dictating the C₁-C₃, C₂-C₄ fold could be applied to other canonical EGF-like domains as well.

However, to provide further support for the conclusion, the same set of experiments performed on t-TM EGF D4 and D5 should be applied to other canonical EGF-like domains (e.g. EGF-like domain 1 of coagulation factor VII and pro-neuregulin-1 EGF-like domain).

4.3 Implication of Findings

The main objective of this thesis is to provide more insights into the interpretation of the protein folding code. Here, attempts were made to shed light on the nature of structural determinants which play a role in maintaining the overall fold of evolutionarily conserved protein domains despite hypervariability in their amino acid sequences.

The concept of structural determinants deviate from the common definition of the protein folding code. In the common definition, the three-dimensional structure of a protein is considered to be dictated by the totality of the amino acid sequence. However, in the case of structural determinants, only certain specific residues fulfill the role of a guide in the folding decision of a protein. This, as mentioned previously, would allow functional diversification to take place on a single protein scaffold.

The results obtained from this study demonstrated that a simple switch in the requirement of hydrophobic interactions to the non-requirement in the folding domain — hypothesized to be mediated by the switch in the physical-chemical properties of the structural determinants — is enough to generate a novel protein fold from a single domain platform. Therefore, a single protein modular unit not only serves as the platform for functional diversification, it also serves as a basis for the evolution of protein structure. This new protein structure could in turn participate in novel functions, thus amplifying the rate of functional diversification. In such a case, an exponential rate of protein evolution could be achieved. This could explain the exponential increase in

the complexity of life forms since the beginning of life approximately 3.8 billion years ago — i.e. For a long time the rate of increase in the complexity of life-forms is very slow. Only in the last 350 million years there was an exponential growth in the multi-cellular eukaryotic lineage in its complexity and diversity [93].



Bibliography



1. Anfinsen, C.B., et al., *The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain*. Proc Natl Acad Sci U S A, 1961. **47**(9): p. 1309-14.
2. Haber, E. and C.B. Anfinsen, *Side-chain interactions governing the pairing of half-cystine residues in ribonuclease*. J Biol Chem, 1962. **237**(6): p. 1839-44.
3. White, F.H., Jr., *Regeneration of native secondary and tertiary structures by air oxidation of reduced ribonuclease*. J Biol Chem, 1961. **236**(5): p. 1353-60.
4. Munoz, V. and L. Serrano, *Intrinsic secondary structure propensities of the amino acids, using statistical phi-psi matrices: comparison with experimental scales*. Proteins, 1994. **20**(4): p. 301-11.
5. Pace, C.N. and J.M. Scholtz, *A helix propensity scale based on experimental studies of peptides and proteins*. Biophys J, 1998. **75**(1): p. 422-7.
6. Minor, D.L., Jr. and P.S. Kim, *Measurement of the beta-sheet-forming propensities of amino acids*. Nature, 1994. **367**(6464): p. 660-3.
7. Dill, K.A., *Dominant forces in protein folding*. Biochemistry, 1990. **29**(31): p. 7133-55.
8. Cordes, M.H., A.R. Davidson, and R.T. Sauer, *Sequence space, folding and protein design*. Curr Opin Struct Biol, 1996. **6**(1): p. 3-10.
9. Kamtekar, S., et al., *Protein design by binary patterning of polar and nonpolar amino acids*. Science, 1993. **262**(5140): p. 1680-5.
10. Minor, D.L., Jr. and P.S. Kim, *Context is a major determinant of beta-sheet propensity*. Nature, 1994. **371**(6494): p. 264-7.
11. Minor, D.L., Jr. and P.S. Kim, *Context-dependent secondary structure formation of a designed protein sequence*. Nature, 1996. **380**(6576): p. 730-4.
12. Mezei, M., *Chameleon sequences in the PDB*. Protein Eng, 1998. **11**(6): p. 411-4.
13. Bulaj, G. and A. Walewska, *Oxidative Folding of Single-stranded Disulfide-rich Peptides*, in *Oxidative Folding of Peptides and Proteins*, J. Buchner and L. Moroder, Editors. 2009, Royal Society of Chemistry. p. 274-96.
14. Dill, K.A., et al., *The protein folding problem*. Annu Rev Biophys, 2008. **37**: p. 289-316.
15. Levinthal, C., *Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois.*, ed. P. Debrunner, J.C.M. Tsibris, and E. Münck. 1969: University of Illinois Press, Urbana.
16. Mayor, U., et al., *Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation*. Proc Natl Acad Sci U S A, 2000. **97**(25): p. 13518-22.
17. Sosnick, T.R., et al., *The barriers in protein folding*. Nat Struct Biol, 1994. **1**(3): p. 149-56.

18. Kim, P.S. and R.L. Baldwin, *Intermediates in the folding reactions of small proteins*. *Annu Rev Biochem*, 1990. **59**: p. 631-60.
19. Ptitsyn, O.B., *Kinetic and equilibrium intermediates in protein folding*. *Protein Eng*, 1994. **7**(5): p. 593-6.
20. Arai, M. and K. Kuwajima, *Rapid formation of a molten globule intermediate in refolding of alpha-lactalbumin*. *Fold Des*, 1996. **1**(4): p. 275-87.
21. Vidugiris, G.J., J.L. Markley, and C.A. Royer, *Evidence for a molten globule-like transition state in protein folding from determination of activation volumes*. *Biochemistry*, 1995. **34**(15): p. 4909-12.
22. Nolting, B., *Analysis of the folding pathway of chymotrypsin inhibitor by correlation of phi-values with inter-residue contacts*. *J Theor Biol*, 1999. **197**(1): p. 113-21.
23. Itzhaki, L.S., D.E. Otzen, and A.R. Fersht, *The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding*. *J Mol Biol*, 1995. **254**(2): p. 260-88.
24. Nolting, B., et al., *The folding pathway of a protein at high resolution from microseconds to seconds*. *Proc Natl Acad Sci U S A*, 1997. **94**(3): p. 826-30.
25. Plotkin, S.S. and J.N. Onuchic, *Understanding protein folding with energy landscape theory. Part I: Basic concepts*. *Q Rev Biophys*, 2002. **35**(2): p. 111-67.
26. Onuchic, J.N., Z. Luthey-Schulten, and P.G. Wolynes, *Theory of protein folding: the energy landscape perspective*. *Annu Rev Phys Chem*, 1997. **48**: p. 545-600.
27. Creighton, T.E., *Disulfide bonds as probes of protein folding pathways*. *Methods Enzymol*, 1986. **131**: p. 83-106.
28. Wedemeyer, W.J., et al., *Disulfide bonds and protein folding*. *Biochemistry*, 2000. **39**(15): p. 4207-16.
29. Creighton, T.E., *Experimental studies of protein folding and unfolding*. *Prog Biophys Mol Biol*, 1978. **33**(3): p. 231-97.
30. Konishi, Y., T. Ooi, and H.A. Scheraga, *Regeneration of ribonuclease A from the reduced protein. Rate-limiting steps*. *Biochemistry*, 1982. **21**(19): p. 4734-40.
31. Weissman, J.S. and P.S. Kim, *A kinetic explanation for the rearrangement pathway of BPTI folding*. *Nat Struct Biol*, 1995. **2**(12): p. 1123-30.
32. Weissman, J.S. and P.S. Kim, *Reexamination of the folding of BPTI: predominance of native intermediates*. *Science*, 1991. **253**(5026): p. 1386-93.
33. Chatrenet, B. and J.Y. Chang, *The disulfide folding pathway of hirudin elucidated by stop/go folding experiments*. *J Biol Chem*, 1993. **268**(28): p. 20988-96.

34. Chang, J.Y., et al., *The disulfide folding pathway of human epidermal growth factor*. J Biol Chem, 1995. **270**(16): p. 9207-16.
35. Arolas, J.L., et al., *Role of kinetic intermediates in the folding of leech carboxypeptidase inhibitor*. J Biol Chem, 2004. **279**(36): p. 37261-70.
36. Chang, J.Y., *The folding pathway of alpha-lactalbumin elucidated by the technique of disulfide scrambling. Isolation of on-pathway and off-pathway intermediates*. J Biol Chem, 2002. **277**(1): p. 120-6.
37. Konishi, Y., T. Ooi, and H.A. Scheraga, *Regeneration of RNase A from the reduced protein: models of regeneration pathways*. Proc Natl Acad Sci U S A, 1982. **79**(18): p. 5734-8.
38. Arolas, J.L., et al., *Folding of small disulfide-rich proteins: clarifying the puzzle*. Trends Biochem Sci, 2006. **31**(5): p. 292-301.
39. Cemazar, M., et al., *Oxidative folding intermediates with nonnative disulfide bridges between adjacent cysteine residues*. Proc Natl Acad Sci U S A, 2003. **100**(10): p. 5754-9.
40. Saez, G., et al., *The production of free radicals during the autoxidation of cysteine and their effect on isolated rat hepatocytes*. Biochim Biophys Acta, 1982. **719**(1): p. 24-31.
41. Kang, T.S., et al., *Effect of C-terminal amidation on folding and disulfide-pairing of alpha-conotoxin Iml*. Angew Chem Int Ed Engl, 2005. **44**(39): p. 6333-7.
42. Savage, C.R., Jr., T. Inagami, and S. Cohen, *The primary structure of epidermal growth factor*. J Biol Chem, 1972. **247**(23): p. 7612-21.
43. Sawmynaden, K., et al., *Structural insights into microneme protein assembly reveal a new mode of EGF domain recognition*. EMBO Rep, 2008. **9**(11): p. 1149-55.
44. Schnepf, B., et al., *EGF domain swap converts a drosophila EGF receptor activator into an inhibitor*. Genes Dev, 1998. **12**(7): p. 908-13.
45. Stenflo, J., Y. Stenberg, and A. Muranyi, *Calcium-binding EGF-like modules in coagulation proteinases: function of the calcium ion in module interactions*. Biochim Biophys Acta, 2000. **1477**(1-2): p. 51-63.
46. Bailey, K., et al., *Action of thrombin in the clotting of fibrinogen*. Nature, 1951. **167**(4241): p. 233-4.
47. Lorand, L. and W.R. Middlebrook, *The action of thrombin on fibrinogen*. Biochem J, 1952. **52**(2): p. 196-9.
48. Medved, L.V., et al., *Localization of a fibrin polymerization site complementary to Gly-His-Arg sequence*. FEBS Lett, 1993. **320**(3): p. 239-42.
49. Olexa, S.A. and A.Z. Budzynski, *Localization of a fibrin polymerization site*. J Biol Chem, 1981. **256**(7): p. 3544-9.
50. Pratt, K.P., et al., *The primary fibrin polymerization pocket: three-dimensional structure of a 30-kDa C-terminal gamma chain fragment*

- complexed with the peptide Gly-Pro-Arg-Pro*. Proc Natl Acad Sci U S A, 1997. **94**(14): p. 7176-81.
51. Takagi, T. and R.F. Doolittle, *Amino acid sequence studies on factor XIII and the peptide released during its activation by thrombin*. Biochemistry, 1974. **13**(4): p. 750-6.
 52. Pisano, J.J., J.S. Finlayson, and M.P. Peyton, *Cross-link in fibrin polymerized by factor 13: epsilon-(gamma-glutamyl)lysine*. Science, 1968. **160**(830): p. 892-3.
 53. Oliver, J.A., et al., *Thrombin activates factor XI on activated platelets in the absence of factor XII*. Arterioscler Thromb Vasc Biol, 1999. **19**(1): p. 170-7.
 54. Fujikawa, K., et al., *The mechanism of activation of bovine factor IX (Christmas factor) by bovine factor XIa (activated plasma thromboplastin antecedent)*. Biochemistry, 1974. **13**(22): p. 4508-16.
 55. Nogami, K., et al., *Factor VIII C2 domain contains the thrombin-binding site responsible for thrombin-catalyzed cleavage at Arg1689*. J Biol Chem, 2000. **275**(33): p. 25774-80.
 56. Duffy, E.J., et al., *Binding of factor VIIIa and factor VIII to factor IXa on phospholipid vesicles*. J Biol Chem, 1992. **267**(24): p. 17006-11.
 57. Duffy, E.J. and P. Lollar, *Intrinsic pathway activation of factor X and its activation peptide-deficient derivative, factor Xdes-143-191*. J Biol Chem, 1992. **267**(11): p. 7821-7.
 58. Suzuki, K., B. Dahlback, and J. Stenflo, *Thrombin-catalyzed activation of human coagulation factor V*. J Biol Chem, 1982. **257**(11): p. 6556-64.
 59. Pryzdial, E.L. and K.G. Mann, *The association of coagulation factor Xa and factor Va*. J Biol Chem, 1991. **266**(14): p. 8969-77.
 60. Esmon, C.T., N.L. Esmon, and K.W. Harris, *Complex formation between thrombin and thrombomodulin inhibits both thrombin-catalyzed fibrin formation and factor V activation*. J Biol Chem, 1982. **257**(14): p. 7944-7.
 61. Esmon, C.T. and W.G. Owen, *Identification of an endothelial cell cofactor for thrombin-catalyzed activation of protein C*. Proc Natl Acad Sci U S A, 1981. **78**(4): p. 2249-52.
 62. Bukys, M.A., et al., *The structural integrity of anion binding exosite I of thrombin is required and sufficient for timely cleavage and activation of factor V and factor VIII*. J Biol Chem, 2006. **281**(27): p. 18569-80.
 63. Dharmawardana, K.R., S.T. Olson, and P.E. Bock, *Role of regulatory exosite I in binding of thrombin to human factor V, factor Va, factor Va subunits, and activation fragments*. J Biol Chem, 1999. **274**(26): p. 18635-43.
 64. Jakubowski, H.V. and W.G. Owen, *Macromolecular specificity determinants on thrombin for fibrinogen and thrombomodulin*. J Biol Chem, 1989. **264**(19): p. 11117-21.
 65. Suzuki, K. and J. Nishioka, *A thrombin-based peptide corresponding to the sequence of the thrombomodulin-binding site blocks the procoagulant activities of thrombin*. J Biol Chem, 1991. **266**(28): p. 18498-501.

66. Kisiel, W., et al., *Anticoagulant properties of bovine plasma protein C following activation by thrombin*. *Biochemistry*, 1977. **16**(26): p. 5824-31.
67. Dahlback, B., *Factor V and protein S as cofactors to activated protein C*. *Haematologica*, 1997. **82**(1): p. 91-5.
68. Dahlback, B. and B.O. Villoutreix, *The anticoagulant protein C pathway*. *FEBS Lett*, 2005. **579**(15): p. 3310-6.
69. Lu, D., et al., *Comparison of activated protein C/protein S-mediated inactivation of human factor VIII and factor V*. *Blood*, 1996. **87**(11): p. 4708-17.
70. Kunz, G., et al., *Naturally occurring mutations in the thrombomodulin gene leading to impaired expression and function*. *Blood*, 2002. **99**(10): p. 3646-53.
71. Isermann, B., et al., *The thrombomodulin-protein C system is essential for the maintenance of pregnancy*. *Nat Med*, 2003. **9**(3): p. 331-7.
72. White, C.E., et al., *Large-scale expression, purification and characterization of small fragments of thrombomodulin: the roles of the sixth domain and of methionine 388*. *Protein Eng*, 1995. **8**(11): p. 1177-87.
73. Meininger, D.P., M.J. Hunter, and E.A. Komives, *Synthesis, activity, and preliminary structure of the fourth EGF-like domain of thrombomodulin*. *Protein Sci*, 1995. **4**(9): p. 1683-95.
74. Kurosawa, S., et al., *A 10-kDa cyanogen bromide fragment from the epidermal growth factor homology domain of rabbit thrombomodulin contains the primary thrombin binding site*. *J Biol Chem*, 1988. **263**(13): p. 5993-6.
75. Fuentes-Prior, P., et al., *Structural basis for the anticoagulant activity of the thrombin-thrombomodulin complex*. *Nature*, 2000. **404**(6777): p. 518-25.
76. Sampoli Benitez, B.A., et al., *Structure of the fifth EGF-like domain of thrombomodulin: An EGF-like domain with a novel disulfide-bonding pattern*. *J Mol Biol*, 1997. **273**(4): p. 913-26.
77. White, C.E., et al., *The fifth epidermal growth factor-like domain of thrombomodulin does not have an epidermal growth factor-like disulfide bonding pattern*. *Proc Natl Acad Sci U S A*, 1996. **93**(19): p. 10177-82.
78. Hunter, M.J. and E.A. Komives, *Thrombin-binding affinities of different disulfide-bonded isomers of the fifth EGF-like domain of thrombomodulin*. *Protein Sci*, 1995. **4**(10): p. 2129-37.
79. Wood, M.J., B.A. Sampoli Benitez, and E.A. Komives, *Solution structure of the smallest cofactor-active fragment of thrombomodulin*. *Nat Struct Biol*, 2000. **7**(3): p. 200-4.
80. Kaiser, E., et al., *Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides*. *Anal Biochem*, 1970. **34**(2): p. 595-8.
81. Vojkovsky, T., *Detection of secondary amines on solid phase*. *Pept Res*, 1995. **8**(4): p. 236-7.

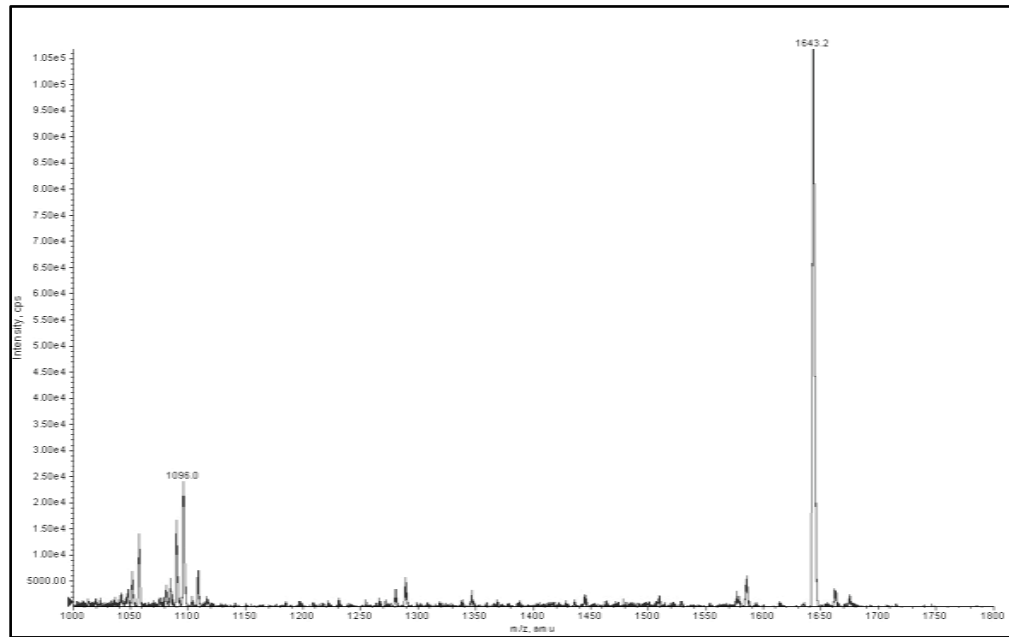
82. Zar, H.J., *Biostatistical Analysis*. 1984, Prentice Hall International: New Jersey. p. 718.
83. Volkin, D.B., H. Mach, and C.R. Middaugh, *Degradative covalent reactions important to protein stability*, in *Protein stability and folding: Theory and practice*, B.A. Shirley, Editor. 1995, Humana Press: Totowa, New Jersey. p. 35-64.
84. Zhang, R.M. and G.H. Snyder, *Factors governing selective formation of specific disulfides in synthetic variants of alpha-conotoxin*. *Biochemistry*, 1991. **30**(47): p. 11343-8.
85. Mason, P.E., et al., *The interaction of guanidinium ions with a model peptide*. *Biophys J*, 2007. **93**(1): p. L04-6.
86. Mande, S.C. and M.E. Sobhia, *Structural characterization of protein-denaturant interactions: crystal structures of hen egg-white lysozyme in complex with DMSO and guanidinium chloride*. *Protein Eng*, 2000. **13**(2): p. 133-41.
87. Monera, O.D., C.M. Kay, and R.S. Hodges, *Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions*. *Protein Sci*, 1994. **3**(11): p. 1984-91.
88. Hao, M.H., et al., *Effects of compact volume and chain stiffness on the conformations of native proteins*. *Proc Natl Acad Sci U S A*, 1992. **89**(14): p. 6614-8.
89. Wouters, M.A., R.A. George, and N.L. Haworth, *"Forbidden" disulfides: their role as redox switches*. *Curr Protein Pept Sci*, 2007. **8**(5): p. 484-95.
90. Shields, P.A. and S.R. Farrah, *Influence of salts on electrostatic interactions between poliovirus and membrane filters*. *Appl Environ Microbiol*, 1983. **45**(2): p. 526-31.
91. Melander, W.R., D. Corradini, and C. Horvath, *Salt-mediated retention of proteins in hydrophobic-interaction chromatography. Application of solvophobic theory*. *J Chromatogr*, 1984. **317**: p. 67-85.
92. Anfinsen, C.B., *The formation and stabilization of protein structure*. *Biochem J*, 1972. **128**(4): p. 737-49.
93. Shah, K.T., *Can there be a limit to the complexity of life?*, in *Life as We Know It*, J. Seckbach, Editor. 2006, Springer: Dordrecht.



Appendix



(A)



(B)

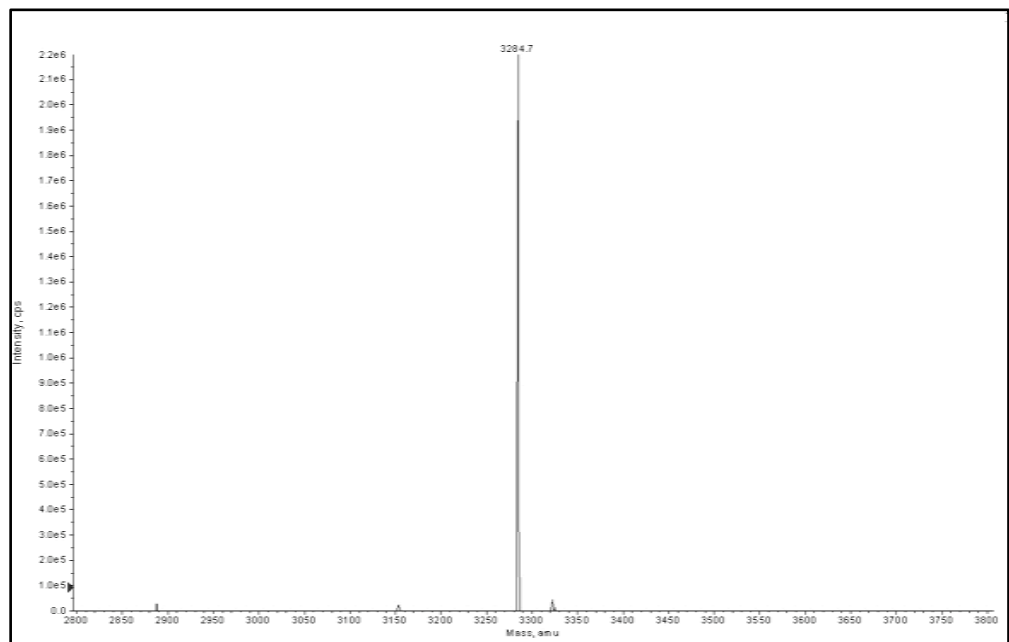
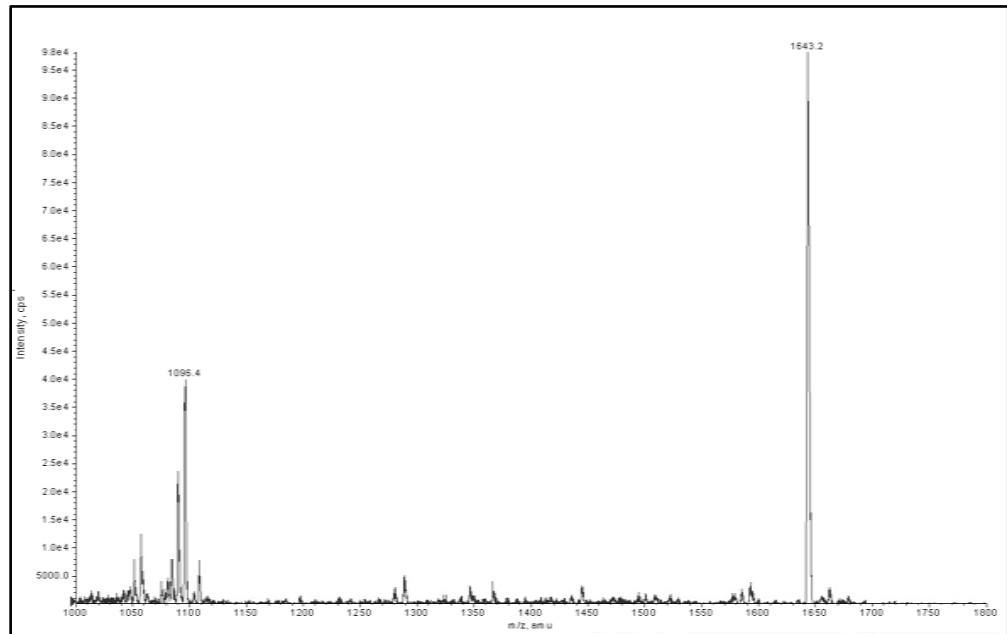


Figure A3.1 ESI-MS mass spectrum of t-TM EGF D4 C₁-C₃, C₂-C₄ (native) isoform. (A) Mass spectrum showing the +3 (m/z: 1096) and +2 (m/z: 1643) charge states of the folded peptide. (B) Reconstructed mass spectrum obtained using the “Peptide Reconstruct” function of the Analyst[®] software.

(A)



(B)

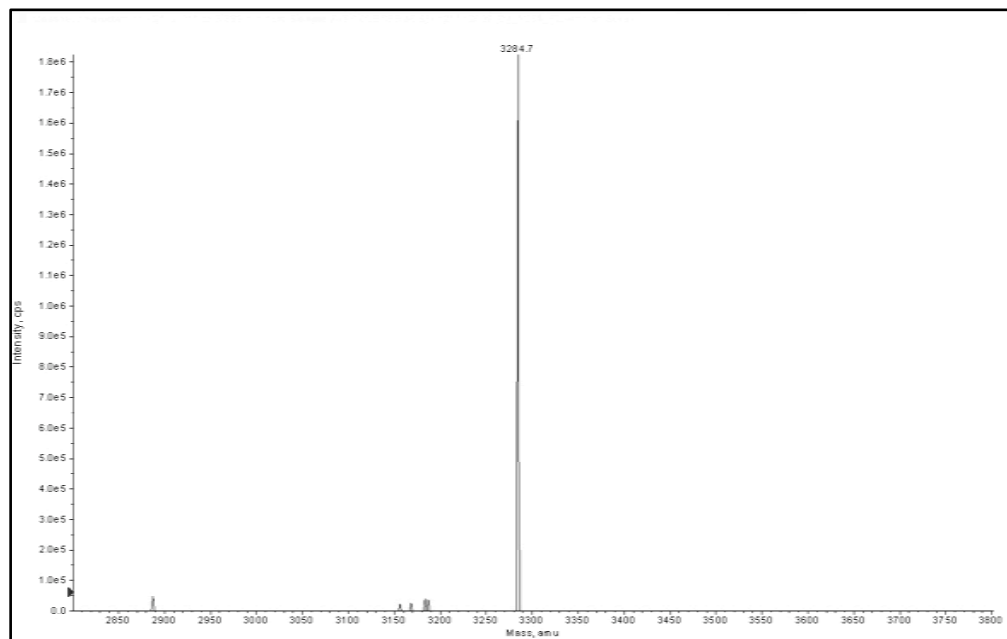
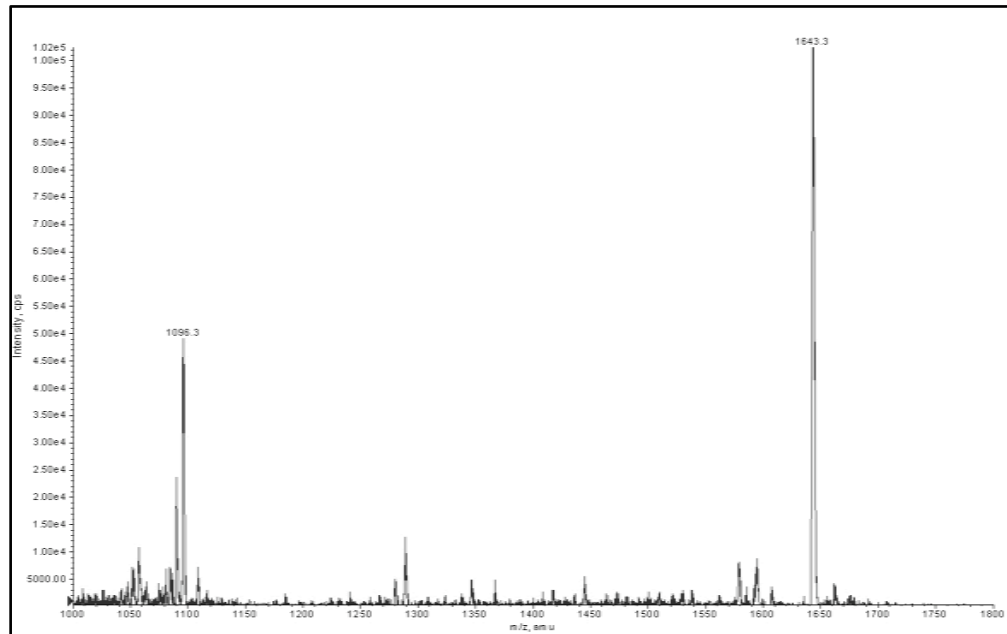


Figure A3.2 ESI-MS mass spectrum of t-TM EGF D4 C₁-C₂, C₃-C₄ isoform. (A) Mass spectrum showing the +3 (m/z: 1096) and +2 (m/z: 1643) charge states of the folded peptide. (B) Reconstructed mass spectrum obtained using the "Peptide Reconstruct" function of the Analyst[®] software.

(A)



(B)

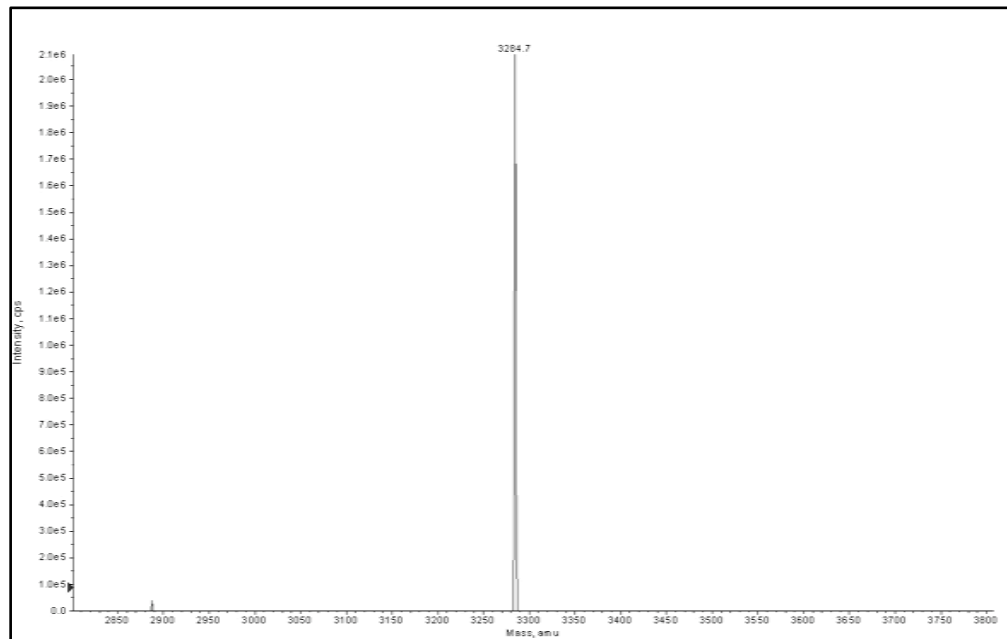


Figure A3.3 ESI-MS mass spectrum of t-TM EGF D4 C₁-C₄, C₂-C₃ isoform. (A) Mass spectrum showing the +3 (m/z: 1096) and +2 (m/z: 1643) charge states of the folded peptide. (B) Reconstructed mass spectrum obtained using the “Peptide Reconstruct” function of the Analyst[®] software.

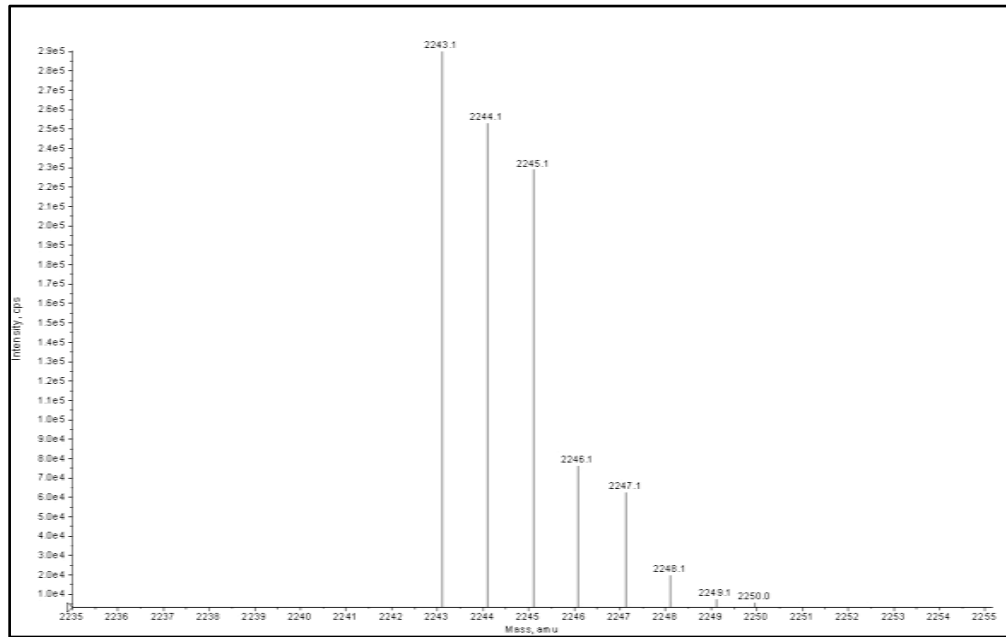


Figure A3.4 Reconstructed ESI-MS mass spectrum of t-TM EGF D5 C₁-C₃, C₂-C₄ isoform. Reconstructed mass spectrum was obtained using the “Peptide Reconstruct” function of the Analyst[®] software. Shown here is the isotopic cluster corresponding to the +1 charge state of the folded peptide. The difference of 1 Da between one isotopic mass to the next confirmed the assignment of the +1 charge state to the isotopic cluster.

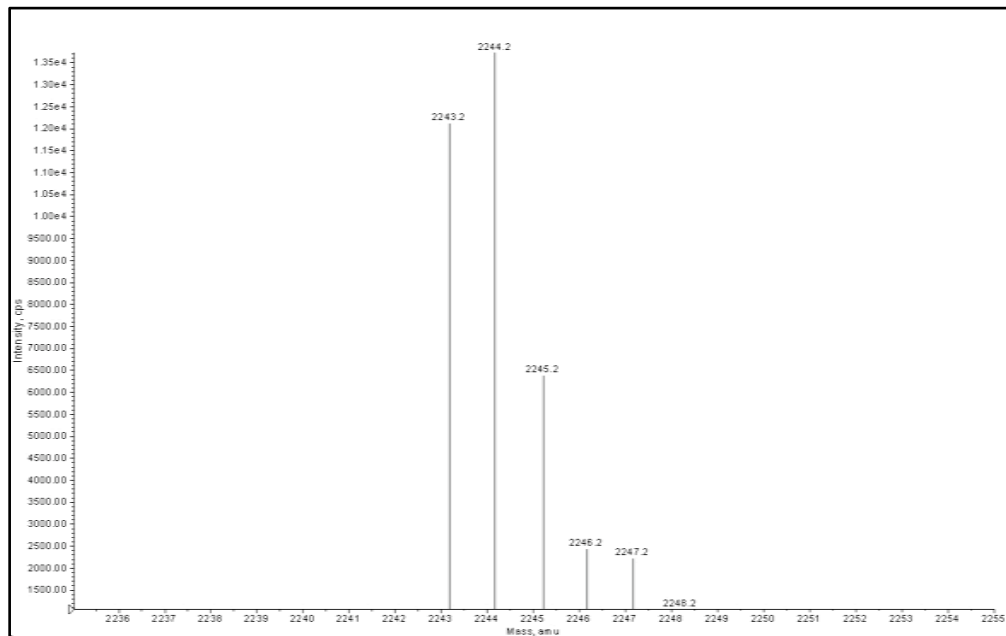


Figure A3.5 Reconstructed ESI-MS mass spectrum of t-TM EGF D5 C₁-C₂, C₃-C₄ (native) isoform. Reconstructed mass spectrum was obtained using the “Peptide Reconstruct” function of the Analyst[®] software. Shown here is the isotopic cluster corresponding to the +1 charge state of the folded peptide. The difference of 1 Da between one isotopic mass to the next confirmed the assignment of the +1 charge state to the isotopic cluster.

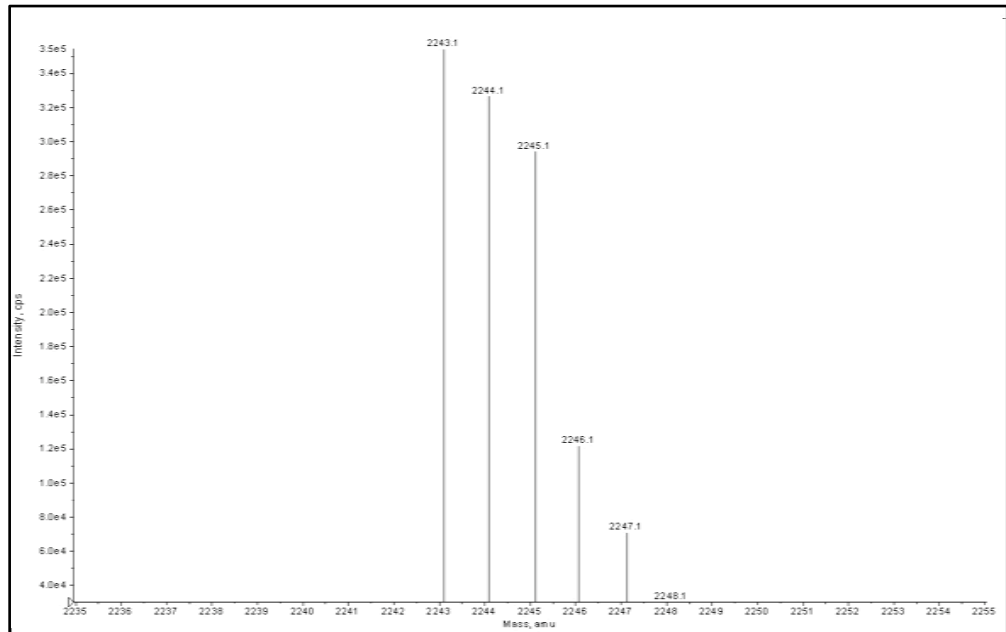


Figure A3.6 Reconstructed ESI-MS mass spectrum of t-TM EGF D5 C₁-C₄, C₂-C₃ isoform. Reconstructed mass spectrum was obtained using the “Peptide Reconstruct” function of the Analyst[®] software. Shown here is the isotopic cluster corresponding to the +1 charge state of the folded peptide. The difference of 1 Da between one isotopic mass to the next confirmed the assignment of the +1 charge state to the isotopic cluster.

Table A3.1 Student's *t*-test (independent samples) for *t*-TM EGF D4 structural isoforms derived from air oxidation and redox reagent-mediated oxidation

t-TM EGF D4: Air Oxidation versus Redox Buffer System									
Structural Isoform	Arcsine Transformed Values				Test statistics				
	Air Oxidation		Redox Buffer System		Mean Difference	Standard Error of Difference	t-statistic (t*)	Degrees of freedom	Probability (p) ^a (one-tailed)
	Mean	Standard Deviation	Mean	Standard Deviation					
C ₁ -C ₃ , C ₂ -C ₄ (Native)	55.26	0.42	56.21	0.35	0.95	0.32	3.00	4	0.02
C ₁ -C ₂ , C ₃ -C ₄	27.36	0.47	25.81	0.42	-1.55	0.36	-4.26	4	0.01
C ₁ -C ₄ , C ₂ -C ₃	19.68	0.56	20.23	0.24	0.55	0.35	1.57	4	0.10

^a Statistical significance when p-value is smaller than 0.05.

Table A3.2 Student's *t*-test (independent samples) for *t*-TM EGF D5 structural isoforms derived from air oxidation and redox reagent-mediated oxidation

t-TM EGF D5: Air Oxidation versus Redox Buffer System									
Structural Isoform	Arcsine Transformed Values				Test statistics				
	Air Oxidation		Redox Buffer System		Mean Difference	Standard Error of Difference	t-statistic (t*)	Degrees of freedom	Probability (p) ^a (one-tailed)
	Mean	Standard Deviation	Mean	Standard Deviation					
C ₁ -C ₃ , C ₂ -C ₄	26.97	0.24	26.82	0.07	-0.14	0.15	-0.98	4	0.19
C ₁ -C ₂ , C ₃ -C ₄ (Native)	51.00	0.38	51.16	0.63	0.16	0.42	0.38	4	0.36
C ₁ -C ₄ , C ₂ -C ₃	25.87	0.22	25.81	0.72	-0.05	0.43	-0.12	4	0.45

^a Statistical significance when p-value is smaller than 0.05.

Table A3.3 Student's *t*-test (independent samples) for *t*-TM EGF D4 structural isoforms derived from air oxidation (with 6 M Gn.HCl) and redox reagent-mediated oxidation (with 6 M Gn.HCl)

t-TM EGF D4: Air Oxidation (with 6 M Gn.HCl) versus Redox Buffer System (with 6 M Gn.HCl)									
Structural Isoform	Arcsine Transformed Values				Test statistics				
	Air Oxidation (+6 M Gn.HCl)		Redox Buffer System (+6 M Gn.HCl)		Mean Difference	Standard Error of Difference	t-statistic (t*)	Degrees of freedom	Probability (p) ^a (one-tailed)
	Mean	Standard Deviation	Mean	Standard Deviation					
C ₁ -C ₃ , C ₂ -C ₄ (Native)	25.46	0.85	34.02	0.60	8.57	0.60	14.23	4	<0.01
C ₁ -C ₂ , C ₃ -C ₄	46.56	0.54	43.44	0.18	-3.12	0.33	-9.51	4	<0.01
C ₁ -C ₄ , C ₂ -C ₃	32.46	0.52	27.56	0.50	-4.89	0.42	-11.77	4	<0.01

^a Statistical significance when p-value is smaller than 0.05.

Table A3.4 Student's *t*-test (independent samples) for *t*-TM EGF D5 structural isoforms derived from air oxidation (with 6 M Gn.HCl) and redox reagent-mediated oxidation (with 6 M Gn.HCl)

t-TM EGF D5: Air Oxidation (with 6 M Gn.HCl) versus Redox Buffer System (with 6 M Gn.HCl)									
Structural Isoform	Arcsine Transformed Values				Test statistics				
	Air Oxidation (+6 M Gn.HCl)		Redox Buffer System (+6 M Gn.HCl)		Mean Difference	Standard Error of Difference	t-statistic (t*)	Degrees of freedom	Probability (p) ^a (one-tailed)
	Mean	Standard Deviation	Mean	Standard Deviation					
C ₁ -C ₃ , C ₂ -C ₄	24.87	0.26	25.92	0.11	1.05	0.16	6.43	4	<0.01
C ₁ -C ₂ , C ₃ -C ₄ (Native)	51.82	0.54	51.63	0.46	-0.20	0.41	-0.48	4	0.33
C ₁ -C ₄ , C ₂ -C ₃	26.93	0.48	25.15	0.46	-0.78	0.38	-2.04	4	0.11

^a Statistical significance when p-value is smaller than 0.05.

Table A3.5 Student's *t*-test (independent samples) for *t*-TM EGF D4 structural isoforms derived from air oxidation and air oxidation (with 6 M Gn.HCl)

t-TM EGF D4: Air Oxidation versus Air Oxidation (with 6 M Gn.HCl)									
Structural Isoform	Arcsine Transformed Values				Test statistics				
	Air Oxidation		Air Oxidation (+6 M Gn.HCl)		Mean Difference	Standard Error of Difference	t-statistic (t*)	Degrees of freedom	Probability (p) ^a (one-tailed)
	Mean	Standard Deviation	Mean	Standard Deviation					
C ₁ -C ₃ , C ₂ -C ₄ (Native)	55.26	0.42	25.45	0.85	-29.81	0.55	-54.35	4	<0.01
C ₁ -C ₂ , C ₃ -C ₄	27.36	0.47	46.56	0.54	19.20	0.41	46.40	4	<0.01
C ₁ -C ₄ , C ₂ -C ₃	19.68	0.56	32.46	0.52	12.78	0.44	28.85	4	<0.01

^a Statistical significance when p-value is smaller than 0.05.

Table A3.6 Student's *t*-test (independent samples) for *t*-TM EGF D4 structural isoforms derived from redox reagent-mediated oxidation and redox reagent-mediated oxidation (with 6 M Gn.HCl)

t-TM EGF D4: Redox Buffer System versus Redox Buffer System (with 6 M Gn.HCl)									
Structural Isoform	Arcsine Transformed Values				Test statistics				
	Redox Buffer System		Redox Buffer System (+6 M Gn.HCl)		Mean Difference	Standard Error of Difference	t-statistic (t*)	Degrees of freedom	Probability (p) ^a (one-tailed)
	Mean	Standard Deviation	Mean	Standard Deviation					
C ₁ -C ₃ , C ₂ -C ₄ (Native)	56.21	0.35	34.02	0.60	-22.19	0.40	-55.03	4	<0.01
C ₁ -C ₂ , C ₃ -C ₄	25.81	0.42	43.44	0.18	17.63	0.26	67.42	4	<0.01
C ₁ -C ₄ , C ₂ -C ₃	20.23	0.24	27.56	0.50	7.33	0.32	22.99	4	<0.01

^a Statistical significance when p-value is smaller than 0.05.

Table A3.7 Student's t-test (independent samples) for t-TM EGF D5 structural isoforms derived from air oxidation and air oxidation (with 6 M Gn.HCl)

t-TM EGF D5: Air Oxidation versus Air Oxidation (with 6 M Gn.HCl)									
Structural Isoform	Arcsine Transformed Values				Test statistics				
	Air Oxidation		Air Oxidation (+6 M Gn.HCl)		Mean Difference	Standard Error of Difference	t-statistic (t*)	Degrees of freedom	Probability (p) ^a (one-tailed)
	Mean	Standard Deviation	Mean	Standard Deviation					
C ₁ -C ₃ , C ₂ -C ₄	26.97	0.24	24.87	0.26	-2.10	0.21	-10.20	4	<0.01
C ₁ -C ₂ , C ₃ -C ₄ (Native)	51.00	0.38	51.82	0.54	0.82	0.38	2.16	4	0.05
C ₁ -C ₄ , C ₂ -C ₃	25.87	0.22	26.93	0.48	1.07	0.30	3.51	4	0.01

^a Statistical significance when p-value is smaller than 0.05.

Table A3.8 Student's t-test (independent samples) for t-TM EGF D5 structural isoforms derived from redox reagent-mediated oxidation and redox reagent-mediated oxidation (with 6 M Gn.HCl)

t-TM EGF D5: Redox Buffer System versus Redox Buffer System (with 6 M Gn.HCl)									
Structural Isoform	Arcsine Transformed Values				Test statistics				
	Redox Buffer System		Redox Buffer System (+6 M Gn.HCl)		Mean Difference	Standard Error of Difference	t-statistic (t*)	Degrees of freedom	Probability (p) ^a (one-tailed)
	Mean	Standard Deviation	Mean	Standard Deviation					
C ₁ -C ₃ , C ₂ -C ₄	26.82	0.07	25.92	0.11	-0.90	0.08	-11.53	4	<0.01
C ₁ -C ₂ , C ₃ -C ₄ (Native)	51.16	0.63	51.63	0.46	0.46	0.45	1.04	4	0.18
C ₁ -C ₄ , C ₂ -C ₃	25.81	0.72	26.15	0.46	0.34	0.49	0.69	4	0.26

^a Statistical significance when p-value is smaller than 0.05.

Table A3.9 Student's *t*-test (independent samples) for *t*-TM EGF D4 structural isoforms derived from redox reagent-mediated oxidation and redox reagent-mediated oxidation (with 0.5 M NaCl)

t-TM EGF D4: Redox Buffer System versus Redox Buffer System (with 0.5 M NaCl)									
Structural Isoform	Arcsine Transformed Values				Test statistics				
	Redox Buffer System		Redox Buffer System (+0.5 M NaCl)		Mean Difference	Standard Error of Difference	t-statistic (t*)	Degrees of freedom	Probability (p) ^a (one-tailed)
	Mean	Standard Deviation	Mean	Standard Deviation					
C ₁ -C ₃ , C ₂ -C ₄ (Native)	56.21	0.35	59.78	0.57	3.56	0.39	9.19	4	<0.01
C ₁ -C ₂ , C ₃ -C ₄	25.81	0.42	24.48	0.34	-1.33	0.31	-4.29	4	0.01
C ₁ -C ₄ , C ₂ -C ₃	20.23	0.24	16.60	0.51	-3.63	0.32	-11.17	4	<0.01

^a Statistical significance when p-value is smaller than 0.05.

Table A3.10 Student's *t*-test (independent samples) for *t*-TM EGF D5 structural isoforms derived from redox reagent-mediated oxidation and redox reagent-mediated oxidation (with 0.5 M NaCl)

t-TM EGF D5: Redox Buffer System versus Redox Buffer System (with 0.5 M NaCl)									
Structural Isoform	Arcsine Transformed Values				Test statistics				
	Redox Buffer System		Redox Buffer System (+0.5 M NaCl)		Mean Difference	Standard Error of Difference	t-statistic (t*)	Degrees of freedom	Probability (p) ^a (one-tailed)
	Mean	Standard Deviation	Mean	Standard Deviation					
C ₁ -C ₃ , C ₂ -C ₄	26.82	0.07	28.87	0.29	2.05	0.17	11.77	4	<0.01
C ₁ -C ₂ , C ₃ -C ₄ (Native)	51.16	0.63	47.53	1.02	-3.64	0.69	-5.25	4	<0.01
C ₁ -C ₄ , C ₂ -C ₃	25.81	0.72	28.15	1.36	2.34	0.88	2.64	4	0.03

^a Statistical significance when p-value is smaller than 0.05.