Structural investigation of tumor differentiation factor

Biotechnology and Applied Biochemistry

Urmi Roy,[†] *Izabela Sokolowska,*[†] *Alisa G. Woods, and Costel C. Darie*[*]

Biochemistry & Proteomics Group, Department of Chemistry & Biomolecular Science, Clarkson University, Potsdam, NY, USA

Abstract.

Tumor differentiation factor (TDF) is a 17 kDa protein produced by the pituitary and secreted into the bloodstream, with no definitive function and incomplete characterization. TDF has the following four cysteine (Cys) residues: Cys17, Cys70, Cys97, and Cys98. To understand the function of TDF, we (1) overexpressed and characterized recombinant TDF (rTDF); (2) investigated native, secreted TDF; and (3) assessed potential disulfide connectivities using molecular modeling. Our results from Western blotting (WB) experiments suggest that rTDF is mostly expressed as insoluble, monomeric, and dimeric forms. Mass spectrometry analysis of the overexpressed rTDF identified a peptide that is a part of TDF protein. WB of the native, secreted TDF detected it as a 50 kDa band. In addition, investigation of TDF by molecular modeling suggests that the Cys residues may form disulfide bridges between Cys17–Cys98 and Cys70–Cys17.

Volume 59, Number 6, November/December 2012, Pages 445–450 •
E-mail: cdarie@clarkson.edu

Keywords: tumor differentiation factor, protein expression, mass spectrometry, molecular modeling, disulfide bridges

1. Introduction

Tumor differentiation factor (TDF) is a 17 kDa protein produced by the pituitary and secreted into the bloodstream. TDF induces the differentiation of human breast and prostate cancer cells *in vitro* and *in vivo* [1],[2]. However, TDF does not have any morphological or biochemical differentiation effect on fibroblasts or on kidney, hepatoma, and leukemic lymphocytic cell lines. The biological activity of TDF on breast and prostate cancer cells has not been reproduced by any other growth factors or pituitary hormones. Its secretion from the pituitary into the bloodstream and its differentiation activity on breast and prostate cancer cells suggest that TDF may have an endocrine role [3].

The molecular mechanism through which TDF induces cell differentiation is not known [1–4]. To shed some light into TDF's function, our laboratory isolated and characterized potential TDF receptor (TDF-R) candidates from breast cancer cell lines, nonbreast cancer cell lines, and from fibroblasts [3],[4]. We concluded that TDF-R candidates are members of the 70 kDa heat shock protein family (HSP70). However, further investigation of potential TDF-R candidates and of TDF protein from both natural and recombinant sources is needed.

Characterization of TDF is still in its infancy, and many questions about this protein are still unanswered. For example, it has been demonstrated that TDF is glycosylated by N-linked oligosaccharides [2], but its amino acid sequence contains no glycosylation site. In addition, nothing is known about the role of TDF in the blood and why it is secreted. Therefore, unveiling some of the mysteries that surround TDF could shed some light on the physiological role of TDF and its implications in physiological and pathological states.

When one plans to solve the crystal structure of a protein, initial assessment of that protein using molecular modeling tools is the first step to understanding its structure. This assessment not only helps in predicting the potential structure and posttranslational modifications such as disulfide connectivities [5],[6], but also helps in estimating the rate of success for obtaining high-quality crystals and solving the crystal structure of that protein [7–10]. However, before any of these experiments can be performed, the identity of the overexpressed protein must be revealed. Although Western blotting (WB) is usually an acceptable method for identification of proteins, it is still indirect and its use is based on the assumption that the antibody used is highly specific. Therefore, analytical methods that provide direct evidence for a protein through amino acid sequencing, such as mass spectrometry (MS), are preferred.

Here, we overexpressed, purified, and characterized recombinant TDF (rTDF); compared the rTDF with its native, secreted counterpart; and investigated the disulfide connectivities in TDF using molecular modeling. Using WB, we concluded that rTDF is expressed mostly as an insoluble protein, in monomeric and dimeric form, with a molecular mass similar to its theoretical mass. The identity of rTDF was confirmed by liquid chromatography–tandem MS (LC–MS/MS) when we identified a peptide that was part of TDF. *De novo* sequencing also

Abbreviations: TDF, tumor differentiation factor; rTDF, recombinant TDF; TDF-R, TDF receptor; WB, Western blotting; MS, mass spectrometry; LC–MS/MS, liquid chromatography–tandem mass spectrometry.

[*]Address for correspondence: Costel C. Darie, PhD, Biochemistry & Proteomics Group, Department of Chemistry & Biomolecular Science, Clarkson University, 8 Clarkson Avenue, Potsdam, NY 13699-5810, USA. Tel.: +1 315 268 7763; Fax: +1 315 268 6610; e-mail: cdarie@clarkson.edu.

[†]Urmi Roy and Izabela Sokolowska contributed equally to this work.

Received 30 July 2012; accepted 8 September 2012
DOI: 10.1002/bab.1046
Published online 20 November 2012 in Wiley Online Library (wileyonlinelibrary.com)

confirmed the identity of the peptide. The native, secreted TDF was detected as a 50 kDa band, which is different from the mass of rTDF, suggesting that it is intensely posttranslationally modified. In addition, investigation of TDF by molecular modeling suggests that the cysteine (Cys) residues may form disulfide bridges between Cys17–Cys98 and Cys70–Cys17.

2. Materials and methods

2.1. Materials

All materials were purchased from Sigma–Aldrich (St. Louis, MO, USA), Fisher Scientific (Pittsburgh, PA, USA), or VWR (Radnor, PA, USA), unless otherwise stated. Human sera were purchased from ProMedDx (Norton, MA, USA), and the work was approved by the Institutional Review Board (IRB exempt).

2.2. Construction of *tdf* clone

A pET22b(+) plasmid containing *tdf* gene suited for overexpression of the TDF protein was synthesized by Biomatik Corporation (Cambridge, Canada). The gene sequence was according to Ref. [2] and was tagged with FLAG at the N-terminus and with His at the C-terminus. The selection antibiotic was ampicillin.

2.3. Overexpression of rTDF protein

A tdf-positive colony was picked for rTDF expression and grown overnight in Luria Broth (LB) medium in a small tube and then transferred into a larger vessel and grown until optical density 550 nm was 0.2–0.5. The bacterial strain used was Rosetta (DE3)pLysS competent cells (Novagen, Madison, WI, USA). The bacteria were then induced to produce rTDF protein using 1 mM isopropyl-beta-D-1-thiogalactopyranoside (IPTG). The bacteria were allowed to produce rTDF for 2 H and then pelleted using a centrifuge with JA-20 rotor, 12,000*g* and 4°C temperature. The disrupted bacteria were then fractionated into soluble and insoluble fractions. Aliquots were taken at each time point for further analysis by SDS-PAGE and Coomassie Brilliant Blue (CBB) staining or by SDS-PAGE and WB using anti-TDF antibodies.

2.4. Purification of rTDF

The bacterial pellet was solubilized using solubilization buffer (8 M urea, 100 mM NaH_2PO_4, 2 mM imidazole, protease inhibitor, pH 8.0) and rTDF was purified using nickel–nitriloacetic acid (Ni–NTA) resin. Briefly, lysates were incubated with Ni–NTA resin for 2 H, loaded on a column, washed with washing buffer (100 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole), and eluted with elution buffer (100 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole). All samples were analyzed by SDS-PAGE and CBB staining or by SDS-PAGE and WB using anti-TDF antibodies.

2.5. SDS-PAGE and WB

SDS-PAGE and WB were performed as described in Ref. [11].

2.6. Mass spectrometry

Analysis of the rTDF was performed by LC–MS/MS using a NanoAcquity UPLC coupled with a QTOF Micro mass spectrometry system (both from Waters Corporation, Milford, MA, USA), as described in Ref. [12], with certain modifications [13]. Data processing and the database search were performed as described in Ref. [14].

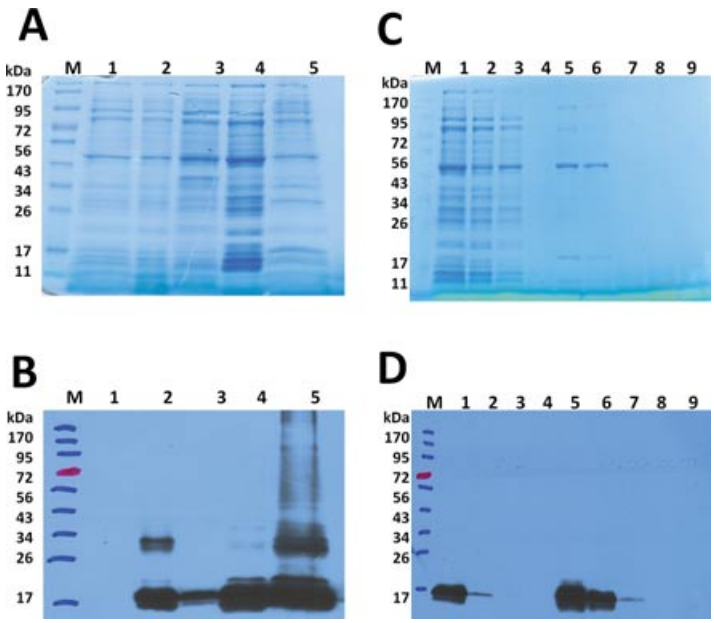

Fig. 1. Overexpression and purification of rTDF. The rTDF was tested for overexpression by Coomassie (A) and WB using anti-TDF antibodies (B). In (A) and (B), the gel lanes contained: cell lysate before IPTG induction (1), cell lysate after 2 H of IPTG induction (2), the soluble cell lysate (3), the insoluble cell lysate (4), and the pellet (5). The rTDF was then purified using Ni column and the purification was tested by Coomassie (C) and WB using anti-TDF antibodies (D). In (C) and (D), the gel lanes contained: the cell lysate before Ni column (1), the cell lysate after Ni column (2), the washing steps (3–4), and elution steps (5–9). The molecular weight markers are indicated.

2.7. Production of anti-TDF antibodies

Anti-TDF antibodies were custom-made by Creative Biolabs (Shirley, NY, USA) using TDF peptide P1 (TDF-P1) as an immunogen. TDF-P1 is a peptide that is part of the open reading frame of TDF and has the sequence NH2-RESQGTRVGQALSFLCKGTA-COOH. TDF-P1 peptide was synthesized by standard peptide synthesis, coupled to Keyhole limpet hemocyanin and used for immunization of rabbits. The correct sequence of TDF-P1 was confirmed by MS and the specificity of the antibody was confirmed by preincubation of anti-TDF antibodies with its peptide antigen in an inhibition assay. The antibodies were purified using Protein A sepharose (Sigma) and concentrated to 1 mg/mL. Working concentration for WB was at a dilution of 1:500.

2.8. Analysis of TDF protein by molecular modeling

A theoretical model of protein structure was predicted based on the amino acid sequence MRESQGTRVGQALSFLCKGTAVQDT-HTHTHTHTHTHTHTNTIQTQNMKHFLYGPFQSLKFTPYFWHHHICP-MIILINAMMLIQLLLEGPDHQYSLECCIFLTFRQVWQ, and using the

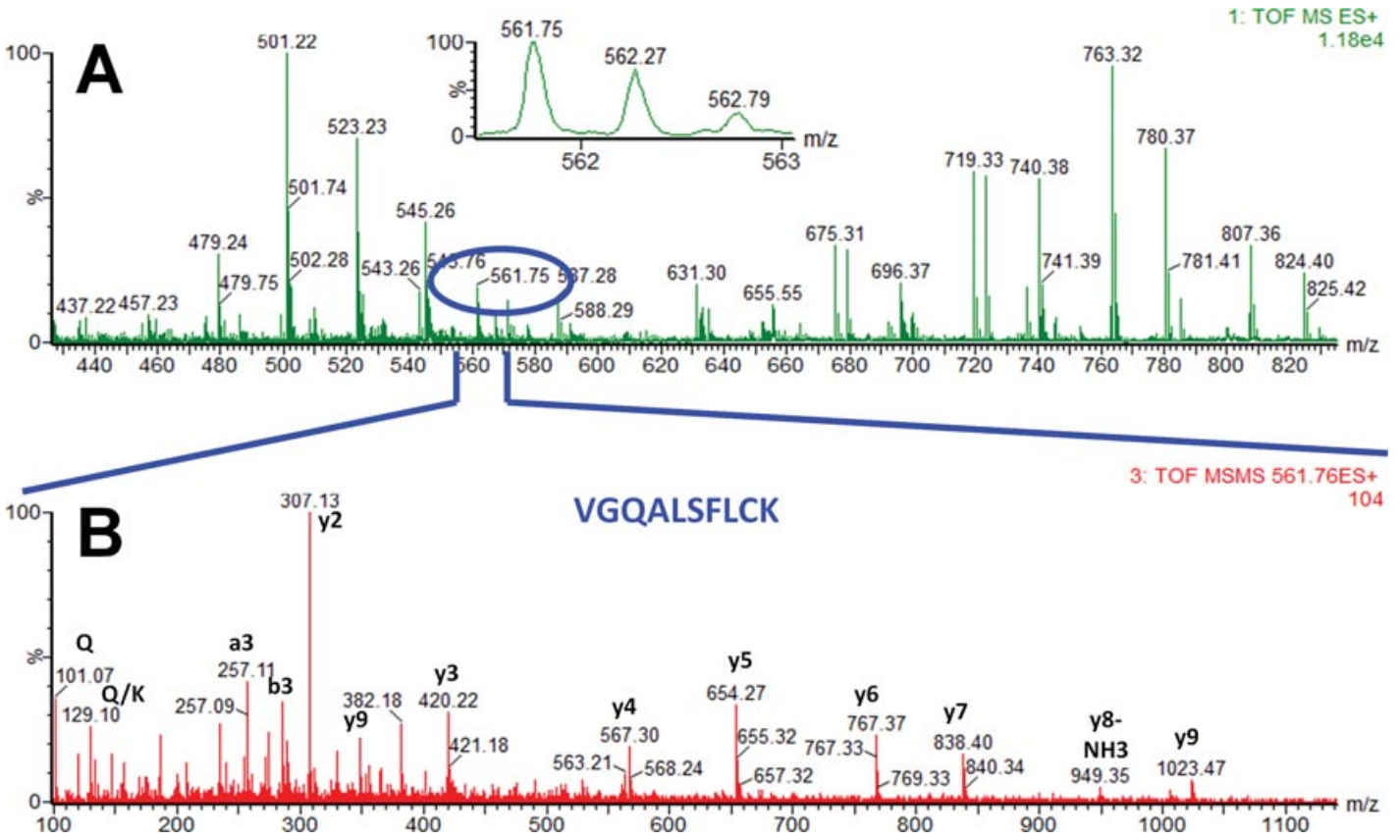

Fig. 2. LC–MS/MS analysis of rTDF. The gel band corresponding to rTDF was excised and digested by trypsin and the resulting peptides mixture was analyzed by LC–MS/MS. (A) A precursor ion with m/z of 561.75 (2+) and shown expanded for determination of the charge state was selected for fragmentation by MS/MS. (B) The fragmentation of the precursor ion though MS/MS produced a series of b and y ions. Data processing and analysis led to the identification of a peptide with the sequence VGQALSFLCK that was part of TDF protein.

I-TASSER server. I-Tasser is an automated protein structure prediction server that produces models using "multiple threading alignments" and "iterative structural assembly simulations" [15],[16].

3. Results and discussion

3.1. rTDF overexpressed mostly as an insoluble monomeric protein

In our initial experiments, we tested the expression of rTDF. We collected different fractions before and after induction of the bacterial cells, as well as after fractionation of the bacterial lysates in the soluble and insoluble fractions. We then tested the expression of rTDF using both SDS-PAGE and CBB or SDS-PAGE and WB. A summary of such an experiment is shown in Figs. 1A (SDS-PAGE and CBB) and 1B (SDS-PAGE and CBB). As observed, rTDF is detected by WB in the whole cell lysate and in the insoluble lysate, as well as in the pellet (Fig. 1B, lanes 2, 4, and 5). Only a small portion of rTDF could be detected in the soluble fraction of the cell lysate (Fig. 1B, lane 3). However, no intensely stained band could be observed in the CBB-stained SDS-PAGE gel (Fig. 1A). Interestingly, although rTDF was mostly detected as a 17 kDa monomeric protein (Fig. 1B, lanes 2, 4, and 5), a small fraction of it was detected as a 32–34 kDa band (Fig. 1B, lanes 2 and 5). Although this band could be a legitimate dimeric rTDF protein, this could also be a separation artifact. Taken together, these data suggest that although there is no visible difference between uninduced and induced samples when analyzed by SDS-PAGE and CBB (Fig. 1A), it is clear that rTDF is expressed (Fig. 1B) in the insoluble fraction.

3.2. rTDF successfully purified as a monomeric protein

No clear visible difference between uninduced and induced samples may suggest that the rTDF expression is too low. However, if a protein expressed at low levels can be successfully purified, as compared to highly overexpressed but difficult to purify proteins (*e.g.*, inclusion bodies), then it is worthwhile to try to purify that protein. Therefore, to investigate whether rTDF can be purified using a fast procedure, we took advantage of the presence of the His tag within the sequence of rTDF. SDS-PAGE analysis of the lysates before and after incubation with the Ni–NTA column, the washing steps, and the eluates is shown in Figs. 1C (CBB staining) and 1D (WB). As observed in WB

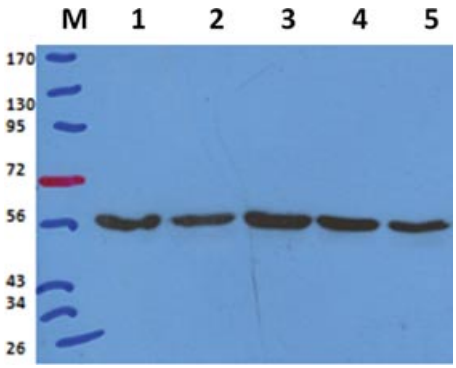

Fig. 3. Analysis of native, secreted TDF by WB in human sera. Equal volumes of sera (5 μL/lane, diluted 1:20) were analyzed by WB using anti-TDF antibodies (1:500 dilution). The sera were from healthy women (all lanes). M represents the marker. The molecular weight markers are indicated.

(Fig. 1D), rTDF was successfully purified using Ni–NTA column. Taken together, these data suggest that although the level of expression of rTDF is low, its purification is fast and effective, as compared to the other expression systems that overexpress proteins but their purification is either difficult (the protein is insoluble) or not possible at all (the proteins do not/cannot be purify/purified using immunoaffinity chromatography or conventional biochemical approaches because of low specificity of the antibodies against the protein of interest or even because of the lack of antibodies).

3.3. Identification by LC–MS/MS analysis of the rTDF of a peptide that is part of TDF

So far, the existence of TDF was demonstrated only through indirect methods such as identification of the *tdf* mRNA through *in situ* hybridization [2] or of its protein product through WB (Fig. 1). To produce direct evidence of the existence of TDF protein, we digested rTDF by trypsin and analyzed the peptide mixture by LC–MS/MS. In our experiments, we identified a doubly charged precursor ion with m/z of 561.75 (Fig. 2A) that was selected for fragmentation in MS/MS (Fig. 2B). Fragmentation of the precursor ion produced a series of b and y ions, and its analysis led to the identification of a peptide with the sequence VGQALSFLCK (with Cys blocked by iodoacetamide) that was part of TDF protein.

3.4. Native, secreted TDF intensely posttranslationally modified

To investigate whether TDF has any posttranslational modification within its amino acid sequence, we separated the human sera by SDS-PAGE and analyzed it by WB. Overall, we concluded that because the molecular mass at which this protein was detected was around 50 kDa, the native, secreted TDF has many posttranslational modifications (Fig. 3). Compared to the molecular mass of rTDF, the posttranslational modifications account for around 30–34 kDa (twice the size of the amino acid sequence of TDF). Overall, these data suggest that the native, secreted TDF is intensely posttranslationally modified.

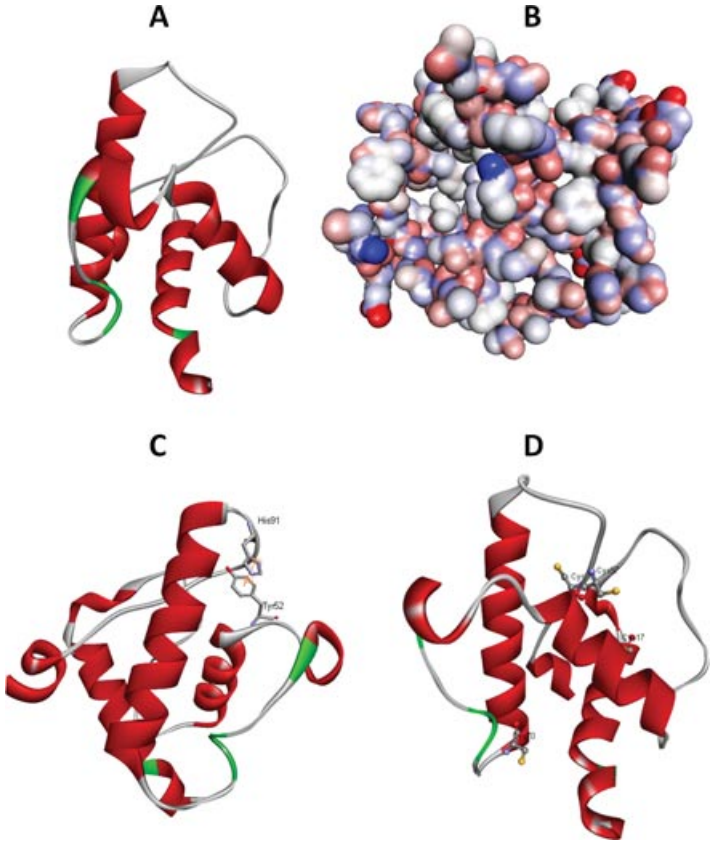

Fig. 4. Possible three-dimensional structure of TDF protein. (A) Ribbon diagram of the possible three-dimensional structure TDF protein. (B) vdW surface depiction of TDF protein, colored by atom charge. (C) pi–pi interaction (orange) in TDF protein. (D) Cysteine residues in TDF protein.

3.5. Insights resulting from investigation of the predicted structure of TDF about the possible disulfide bridges as restriction points within the TDF three-dimensional structure

A theoretical model protein structure was predicted based on the amino acid sequence of TDF and using I-TASSER server [15],[16]. The C-score of this model TDF structure is 4.25, and the template modeling score is 0.27 $\pm$ 0.08, whereas the root-mean-square deviation is relatively high, 14.0 $\pm$ 3.9 Å. A ribbon diagram of the TDF model structure is presented in Fig. 4A, and a van der Waals (vdW) surface profile of the protein is depicted in Fig. 4B. There are one Arg/Trp and three Lys/Phe interacting pairs in this model structure but none of them form energetically significant cation–pi interactions [17]. A pi–pi interaction is observed between Tyr 52 and His 91 (Fig. 4C).

There are four Cys residues (Cys17, Cys70, Cys97, and Cys98) present in TDF protein (Fig. 4D). Neighbors of these Cys residues are depicted in Fig. 5A. Schematic diagrams

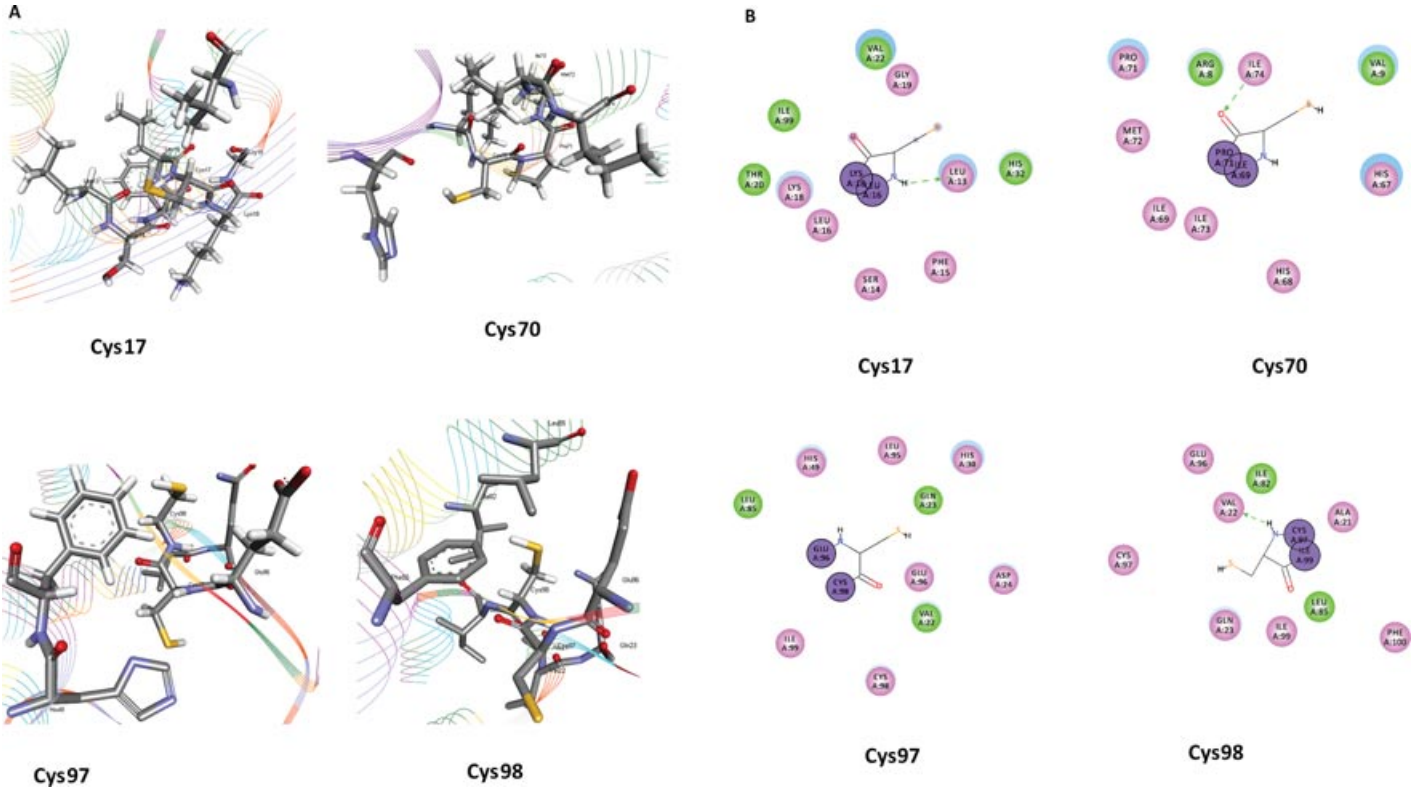

Fig. 5. Cysteine (Cys) residues in TDF. (A) Neighbor residues of Cys in TDF protein. (B) Schematic diagrams showing the interactions between Cys and their neighboring residues in the TDF structure. The pink circles represent the residues participating in hydrogen bonds, as well as in charge or polar interactions. The green circles are residues participating in van der Waals interactions. The magenta circles represent covalently bonded residues. The light blue circle surrounding a given residue/atom denotes the latter's solvent accessible surface. The main chain hydrogen bonds are indicated by the green dashed arrows. These figures were generated using the two-dimensional diagram feature of Accelrys Discovery Studio 3.1 (color by default).

showing the interactions between Cys and amino acid neighbors are displayed in Fig. 5B. Figures 4 and 5 were generated using the Accelrys (San Diego, CA, USA) DS visualizer [18]. The exposed residues of the TDF protein are shown in Fig. 6, where the Cys residues are essentially buried. Figure 6 was originally generated using Deepview/Swiss Pdb Viewer [19] and subsequently was subjected to minor modifications using the Accelrys DS visualizer [18]. Disulfide bonds strongly affect the stability and folding characteristics of a protein and also control the associated enzyme activity. In a protein structure, the half-cystine phase may form disulfide bonds (oxidized), whereas free Cys exists in the reduced form (–SH, they are nondisulfide bridged Cys), and Cys residues participate in ligand binding. Classifications of the Cys residues present in this TDF structure are tabulated in Table 1. This Cys classification was done using the DiANNA web server. Molecular modeling studies using the DiANNA disulfide connectivity server indicate that within TDF protein, the number of disulfide bridges are zero [20]. However, DiANNA connectivity prediction predicts that the probability of predicted connectivity and predicted disulfide bond in this structure are between Cys17–Cys98 (ALSFLCKGTAV–YSLECCIFLTF) and Cys70–Cys 97 (WHHHICPMIIL–QYSLECCIFLT). DiANNA disulfide bonds prediction use trained the neural network and DiANNA to calculate Cys oxidation state (with 76% accuracy) and disulfide bond connectivity (86% accuracy) [21].

Table 1
Cysteine classification in TDF protein

Position of cysteine in TDF	Half cysteine (oxidized form)	Free cysteine (reduced form –SH)	Ligand bound
17	0.184868	0.599401	0.215730
70	0.299266	0.628208	0.072526
97	0.332514	0.496555	0.170931
98	0.345301	0.474789	0.179910

4. Conclusions

We overexpressed and purified rTDF and then analyzed and identified it by LC–MS/MS. We also concluded that native, secreted TDF is intensely posttranslationally modified. In addition, we identified the predicted disulfide bridges that will allow us to use restriction points in the crystal structure of TDF.

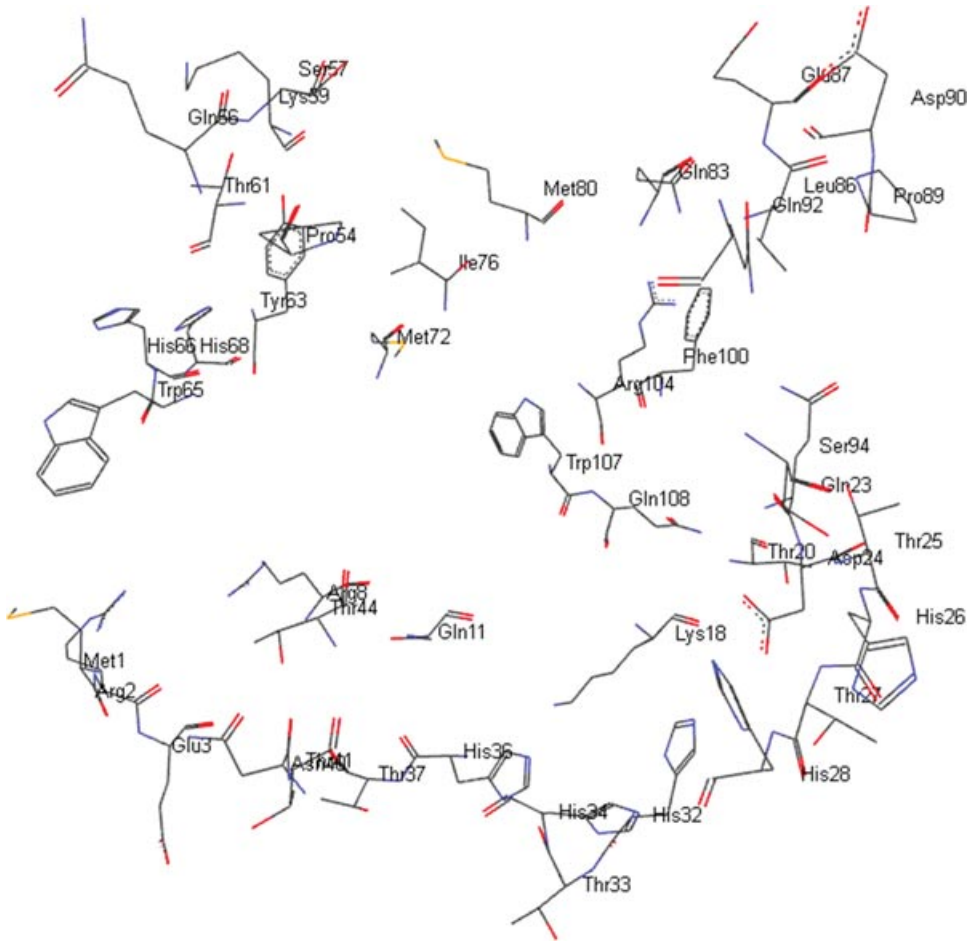

Fig. 6. Exposed residues in model TDF structure.

Acknowledgements

This work was supported in part by the Keep a Breast Foundation (KEABF-375–35054) and by the U.S. Army research office through the Defense University Research Instrumentation Program (DURIP grant #W911NF-11-1-0304).

The authors declare that they have no conflict of interest regarding the manuscript.

References

[1] Platica, M., Chen, H. Z., Ciurea, D., Gil, J., Mandeli, J., and Hollander, V. P. (1992) *Endocrinology* **131**(6), 2573–2580.

[2] Platica, M., Ivan, E., Holland, J. F., Ionescu, A., Chen, S., Mandeli, J., Unger, P. D., and Platica, O. (2004) *Proc. Natl. Acad. Sci. USA* **101**(6), 1560–1565.

[3] Sokolowska, I., Woods, A. G., Gawinowicz, M. A., Roy, U., and Darie, C. C. (2012) *J. Biol. Chem.* **287**(3), 1719–1733.

[4] Sokolowska, I., Woods, A. G., LaFleur, M., Wagner, J., Dorler, J., and Darie, C. C. (2012) In: *Oxidative Stress: Diagnostics and Therapy. ACS Book Series.* Silvana Andreescu, E. and Hepel, M., Eds. **1083**, 341–367, doi: 10.1021/bk-2011-1083.ch013.

[5] Darie, C. C., Litscher, E. S., and Wassarman, P. M. (2008) *Applications of Mass Spectrometry in Life Safety. NATO Science for Peace and Security Series A: Chemistry and Biology*. pp. 23–36, Springer-Verlag, Düsseldorf, Germany.

[6] Darie, C. C., Biniossek, M. L., Gawinowicz, M. A., Milgrom, Y., Thumfart, J. O., Jovine, L., Litscher, E. S., and Wassarman, P. M. (2005) *J. Biol. Chem.* **280**(45), 37585–37598.

[7] Han, L., Monne, M., Okumura, H., Schwend, T., Cherry, A. L., Flot, D., Matsuda, T., and Jovine, L. (2010) *Cell* **143**(3), 404–415.

[8] Jovine, L., Darie, C. C., Litscher, E. S., and Wassarman, P. M. (2005) *Annu. Rev. Biochem.* **74**, 83–114.

[9] Monne, M., Han, L., Schwend, T., Burendahl, S., and Jovine, L. (2008) *Nature* **456**(7222), 653–657.

[10] Monne, M., and Jovine, L. (2011) *Biol. Reprod.* **85**(4), 661–669.

[11] Darie, C. C., Deinhardt, K., Zhang, G., Cardasis, H. S., Chao, M. V., and Neubert, T. A. (2011) *Proteomics* **11**(23), 4514–4528.

[12] Woods, A. G., Sokolowska, I., Yakubu, R., Butkiewicz, M., LaFleur, M., Talbot, C., and Darie, C. C. (2012) In: *Oxidative Stress: Diagnostics and Therapy. ACS Book Series.* Silvana Andreescu, E. and Hepel, M., Eds. **1083**, 369–411, doi: 10.1021/bk-2011-1083.ch012.

[13] Woods, A.G., Sokolowska, I., and Darie, C. C. (2012) *Biochem. Biophys. Res. Commun.* **419**(2), 305–308.

[14] Sokolowska, I., Woods, A. G., Gawinowicz, M. A., Roy, U., and Darie, C. C.. (2012) *FEBS J.* **279**(14), 2579–2594.

[15] Roy, A., Kucukural, A., and Zhang, Y. (2010) *Nat. Protoc.* **5**(4), 725–738.

[16] Zhang, Y. (2008) *BMC Bioinformatics* **9**, 40.

[17] Gallivan, J. P., and Dougherty, D. A. (1999) *Proc. Natl. Acad. Sci. USA* **96**(17), 9459–9464.

[18] Accelrys_Software_Inc. (2012) Discovery Studio Modeling Environment, Release 3.1. Accelrys Software, Inc., San Diego, CA. **1**, 1–1.

[19] Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* **18**(15), 2714–2723.

[20] Ferre, F., and Clote, P. *Nucleic Acids Res.* **33**(Web Server issue), W230–W232.

[21] Ferre, F., and Clote, P. *Nucleic Acids Res.* **34**(Web Server issue), W182–W185.