Mapping the surface areas of *Escherichia coli* glutaredoxin 2 used for its interaction with protein ligands

Christos Karamanos[1], Eleni Poulou-Sidiropoulou[1], Charalampos N. Bompas[1], Haralampos Tzoupis[1], Theodore Tselios[1], Martina Samiotaki[2], and Alexios Vlamis-Gardikas[1]
[1] Department of Chemistry, University of Patras, 26504 Rion, Patras, Greece
[2] Institute for Bioinnovation, Biomedical Sciences Research Center "Alexander Fleming", 16672, Vari, Attica, Greece

Introduction

The thioredoxin (Trx) and glutaredoxin (Grx) enzyme systems reduce disulfides of target protein substrates in the cytosols of all living species including viruses. The Grx system of *Escherichia coli* (*E. coli*) is composed of EcoGrx1-4, glutathione reductase and NADPH [1,2]. EcoGrx2 is an atypical, rather large Grx (215 amino acids, 24,3 kDa) that comprises up to 1 % of total soluble protein in the stationary phase of growth and with strong general antioxidant properties. Its functions however have remained largely unknown [1], until work from our lab that used affinity chromatography and mass spectrometry, highlighted the protein interactome of EcoGrx2. The present work applied *in silico* approaches to pinpoint the critical amino acid residues (contact hot spots) and greater surfaces for the interactions of EcoGrx2 with its protein ligands.

Methodology

First, PDB files were extract from the Uniprot protein database (https://www.uniprot.org). Prior to molecular docking, heteroatoms and non-protein residues were removed. Protein complexes along with their predicted interfaces and docking scores were created using the web tool Prism [3]. Dissociation constants and Gibbs free energy bindings were calculated using the web tool Prodigy. All complexes (Prism) were analyzed by Biovia to identify the contact residues of EcoGrx2 with its ligands. For each amino acid in each complex, the number and type of interaction was recorded, followed by statistical analysis of its participation in contact interfaces. The most statistically significant residues by Biovia ("contact hot spots" [4]) were compared to the thermodynamic hot spots identified by the web tool Proton (http://proton.tools.ibg.edu.tr). Energy network analysis (https://bioinf.iiit.ac.in/NAPS) was applied on the contact hot spots to highlight potential interfaces for the interactions of EcoGrx2 with its protein ligands.

Results

1. Docking

Table 1: Apparent equilibrium dissociation constants (K_Ds) and Gibbs free energy changes (ΔGs) for the interaction of monothiol EcoGrx2 and 38 selected ligands. AF corresponds to structures predicted by Alpha Fold. The interactions between EcoGrx2 and ligands correspond to covalent (C, confirmed by DTT elution), strong non-covalent (NC, confirmed by acid elution). Some proteins (NC, C) were detected by both elution types.

#No.	UniProt ID	gene	PDB ID	Docking Score (kcal)	Interface ID	ΔG (kcal mol^{-1})	Kd (M) 25°C	Interaction type
1	P23857	*pspE*	2JTQ	-40,42	3k9rCD	-12	1.5e-09	sNC, C
2	P0ACB2	*hemB*	1B4E	-6,33	3ez4BG	-11,2	6.2e-09	sNC, C
3	P0A9M0	*lon*	6UZ5	-18,07	1q3eAB	-9,2	1.8e-07	sNC, C
4	P77150	*pdxY*	1TD2	-5,37	1t3uAB	-7,9	1.6e-06	sNC, C
5	P60651	*speB*	7LOL	-19,02	1qjgEF	-11,9	1.8e-09	sNC, C
6	P0853	*tnaA*	2C44	-12,6	1viqAB	-7,5	2.9e-06	sNC, C
7	P37747	*glf*	1I8T	-10,98	2z5eAB	-12,3	9.2e-10	sNC, C
8	P0AF93	*ridA*	1QU9	-12,18	1jkxAB	-9,1	2.2e-07	exponential phase sNC
9	P11875	*argS*	5YYN	-15.08	1o7yAB	-10.4	2.6e-08	exponential phase sNC
10	P31120	*glmM*	AF	-11,93	43adCD	-12,1	1.4e-09	exponential phase sNC
11	P0A9J6	*rbsK*	1GQT	-2,29	3e8oAB	-8,0	1.4e-06	exponential phase sNC
12	P07395	*pheT*	6OZ5	-66,96	3mesAB	-13,7	9.5e-11	exponential phase sNC
13	P07004	*proA*	AF	-8,76	1q3eAB	-7,7	2.4e-06	exponential phase sNC
14	P0A6S7	*gpsA*	AF	-17,35	3g33BD	-14,0	5.4e-11	exponential phase sNC
15	P08373	*murB*	1MBB	-20,01	3he3CF	-9,3	1.5e-07	exponential phase sNC
16	P0A9J8	*pheA*	1ECM	-29.32	1txpCD	-14.9	1.2e-11	exponential phase sNC
17	P00561	*thrA*	AF	-8,72	1tilEF	-9,0	2.4e-07	exponential phase sNC
18	P76187	*ydhF*	1OG6	-3,03	3ftnAB	-7,2	5.6e-06	exponential phase sNC
19	P77398	*arnA*	1Z7E	-24,08	3kf8CD	-7,6	2.9e-06	stationary phase sNC
20	P77774	*bamB*	3RPW	-31,62	1yllAB	-13,4	1.5e-10	stationary phase sNC
21	P0A9S1	*fucO*	1RRM	-19,21	1rrmAB	-13,8	7.5e-11	stationary phase sNC
22	P22256	*gabT*	1SFF	-16,6	1n4aAB	-13	2.9e-10	stationary phase sNC
23	P15288	*pepD*	AF	-1,42	2b8nAB	-7,7	2.6e-06	stationary phase sNC
24	P00561	*thrA*	AF	-6,43	2hfsAB	-9,9	5.1e-08	stationary phase sNC
25	P0A729	*yceF*	4JHC	-5,45	1tqjCD	-7,1	6e-06	stationary phase sNC
26	P0A6L2	*dapA*	1DHP	-16.3	3d31AB	-8,6	5e-07	stationary phase C
27	P0CB39	*eptC*	AF	-28,08	3khnAB	-14,3	3e-11	stationary phase C
28	P05042	*fumC*	1FUO	-27,46	1k50BD	-8,2	1e-06	stationary phase C
29	P0AAC8	*iscA*	1R94	-29,5	3nauAB	-8,7	4e-07	stationary phase C
30	P0ACD4	*iscU*	2KQK	-38,87	2b8nAB	-9,1	2.2e-07	stationary phase C
31	P64463	*ydfZ*	AF	-17,69	3fk9AB	-8,3	8.7e-07	stationary phase C
32	P0AFJ1	*yjdM*	AF	-8,53	1t3uAB	-7,8	1.9e-06	stationary phase C
33	P16095	*sdaA*	AF	-14,54	2wgvAB	-14,5	2.5e-11	exponential phase C
34	P39364	*sgcQ*	AF	-24,11	3dh7CD	-9,2	1.9e-07	exponential phase C
35	P27248	*gcvT*	1VLO	-8,17	3i54AB	-10,6	1.6e-08	exponential phase C
36	P11880	*murF*	1GG4	-5,69	1zbhAB	-8,6	5e-07	exponential phase C
37	P05459	*pdxB*	AF	-16,08	3outAB	-6,9	9.4e-06	exponential phase C
38	P0A7Z0	*rpiA*	1KS2	-7,54	3buaAB	-12	1.6e-09	exponential phase C

Monothiol Grx2 was from pdb 1g7o *in silico* mutated to mono thiol

2. The favored residues and surfaces of EcoGrx2 for its interactions with protein ligands

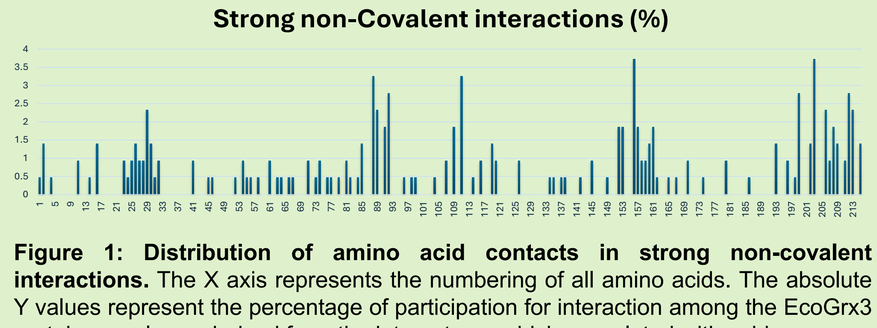

Figure 1: Distribution of amino acid contacts in strong non-covalent interactions. The X axis represents the numbering of all amino acids. The absolute Y values represent the percentage of participation for interaction among the EcoGrx3 protein complexes derived from the interactome which was eluted with acid.

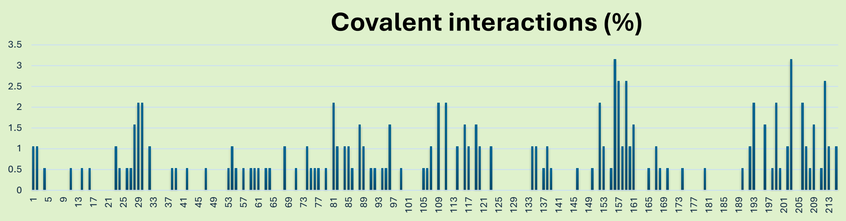

Figure 2: Distribution of amino acid contacts in covalent interactions. The X axis represents the numbering of all amino acids. The absolute Y values represent the percentage of participation among the EcoGrx3 protein complexes derived from the interactome which was eluted with DTT.

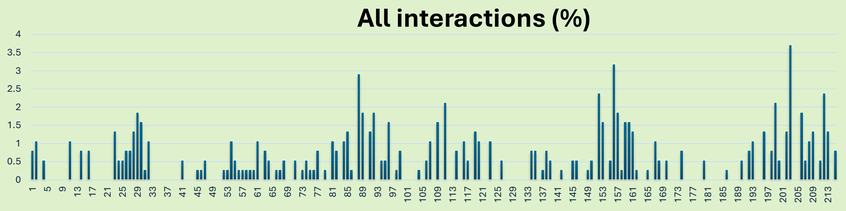

Figure 3: Distribution of amino acid contacts in strong non-covalent and covalent interactions. The X axis represents the numbering of all amino acids. The absolute Y values represent the percentage of participation for interaction among all EcoGrx2 protein complexes.

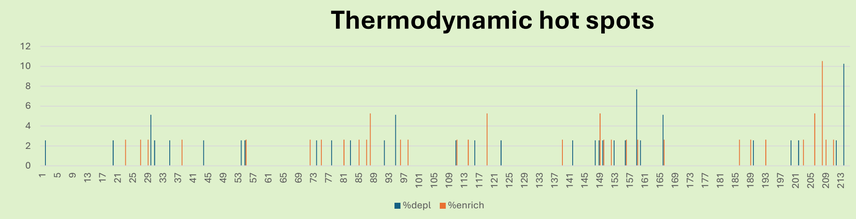

Figure 4: Predicted thermodynamic residues for the interaction of EcoGrx2 with its protein ligands. The X axis represents the numbering of all amino acids. The absolute Y values represent the statistical rate (%) for a potential thermodynamic contact spot.

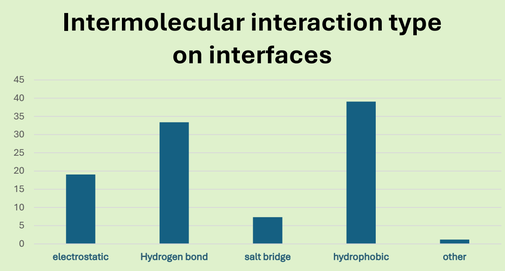

Figure 5: Participation rates of intermolecular interaction types of amino acids on the interfaces in complexes of EcoGrx2 and its protein ligands. Results represent percentages of the types of interactions considering all examined complexes of EcoGrx2.

Calculation of standard deviation Z for each amino acid of EcoGrx2

Poisson approximation

- Interactions per amino acid are not continuous values but counted events
- Interactions are independent events because each complex with EcoGrx2 is unique
- The expected probability of an amino acid participating in an interaction is relatively small: **P = 1 / 215**
- The number of events is large **(379 interactions)**

$$z_i = \frac{\Pi_i - A}{\sqrt{A}}$$

Πi: Observed number of interactions for amino acid i

Expected value: E(X) = λ = A

A = P × 379 = 1/215 × 379 ≈ 1.7

Variance: Var(X) = σ^2 = λ = A

Standard deviation: σ = √λ = √A

EcoGrx2 critical residues with Z > 3.291 (99.9 % confidence) contact hot spots:

His^{29}, Val^{30}, Glu^{88}, Trp^{89}, Lys^{92}, Asp^{111}, Arg^{152}, Lys^{156}, Leu^{157}, Asp^{199}, Lys^{203}, Gln^{206}, Ser^{212}

Table 2: Energy network neighbors for all contact hot spot residues of EcoGrx2. All residues are presented as numbers. "C" stands for covalent and "sNC" for strong non-covalent interactions. The presented Cys^9 is not considered a contact hot spot.

Hot spot	Interaction	Energy neighbors	all DTT	all ACID	common	only DTT	only Acid
Cys^9	Active site	6-13	2-7 20 27-32 112 148-156 152-159 191 195-208 210-214	2-5 20 27-30 84-96 109-114 152-158 195-208 210-214	2-5 20 27-30 112 152-158 195-208 210-214	6, 7 31, 32 148-151 159 191	84-96 109-111 113,114
His^{29}	ACID, DTT	2-5, 27-30					
Val^{30}	DTT	4-7, 28-32, 112					
Glu^{88}	ACID	84-92					
Trp^{89}	ACID	85-93					
Lys^{92}	ACID	88-96					
Asp^{111}	ACID, DTT	109-114					
Arg^{152}	DTT	148-156, 191					
Lys^{156}	ACID, DTT	152-158					
Leu^{157}	DTT	153-159					
Asp^{199}	ACID, DTT	195-203					
Lys^{203}	ACID, DTT	199-207					
Gln^{206}	ACID, DTT	20, 202-208					
Ser^{212}	ACID, DTT	210-214					

Table 3: Parameters for Network Analysis.

Network type:	C-alpha
Lower threshold:	0 Å
Upper threshold:	7 Å
Residue separation:	1
Nodes:	215
Eges:	868
Diameter:	12.00
Radius:	7.00
Average degree:	8.07
Average shortest path:	5.21
Clustering coefficient:	0.53

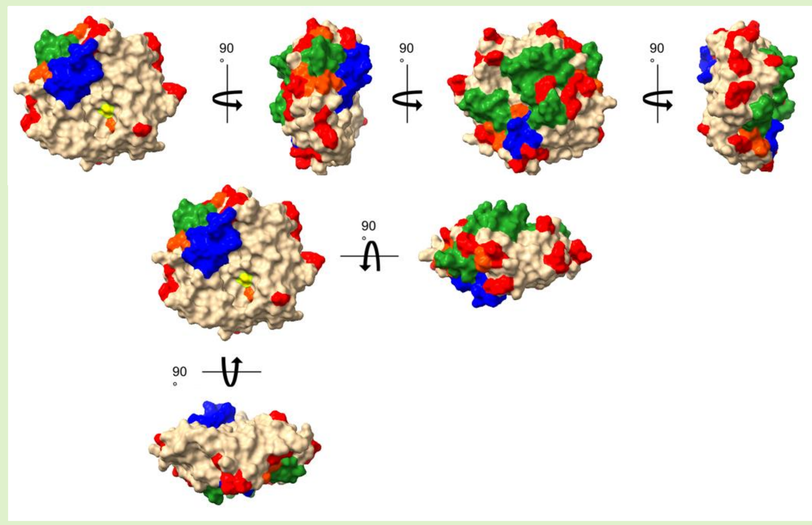

Figure 6: Surface areas of EcoGrx2 involved in the recognition of protein ligands. In green are surfaces participating both in covalent and strong non-covalent interactions, in orange are areas involved in covalent bonding and in blue are regions involved in strong non-covalent interactions. Red residues represent the predicted thermodynamic contact spots (25/54) that are not included within the potential interacting areas.

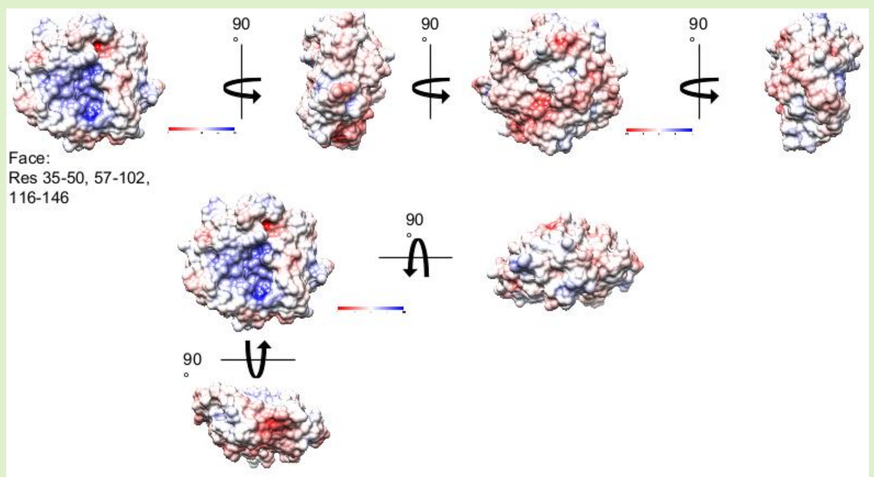

Figure 7: Proposed surfaces based on the charge distribution in the aforementioned areas. Positively charged residues are in red, while negatively charged in blue.

Conclusions

1. Successful molecular docking was performed for 38 verified protein ligands of EcoGrx2.
2. The contact hot spots and their energy neighbors defined three major areas for interactions (mainly hydrophobic and hydrogen bonds) with substrates, all distant from the active site of EcoGrx2.
3. Most of the contact hot spots were common for covalent and strong non-covalent interactions.
4. The statistically favored surface residues (contact hot spots) coincided with 9 thermodynamic hot spots, while 29/54 of the latter were within the herein highlighted contact areas of EcoGrx2.
5. The largely positively charged active site area was devoid of interacting contacts. It is logical to conclude that the area around the active site is reserved for interactions with small negative molecules (e.g. glutathione).

References

1.Vlamis-Gardikas, A. (2008). The multiple functions of the thiol-based electron flow pathways of Escherichia coli: Eternal concepts revisited. *Biochim. Biophys. Acta* 1780(11):1170-200.
2.Ogata F. T., Branco V., Vale F. F., Coppo, L. (2021). Glutaredoxin: Discovery, redox defense and much more. *Redox Biology*, 43, 101975.
3.Tuncbag, N., Gursoy, A., Nussinov, R., Keskin, O. (2011) Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. *Nat. Protoc.* 2011;6:1341–1354.
4.Chasapis, C. T., Makridakis, M., Damdimopoulos, A. E, Zoidakis, J., Lygirou, V., Mavroidis, M., Vlahou, A., Miranda-Vizuete, A., Spyrou, G., Vlamis-Gardikas, A. (2019). Implications of the mitochondrial interactome of mammalian thioredoxin 2 for normal cellular function and disease. *Free Radic. Biol. Med.* 137:59-73.

Acknowledgements

This project was funded by the Hellenic Foundation for Research and Innovation (HFRI) under the "2nd Call for H.F.R.I. Research Projects to Support Faculty Members & Researchers" (Project Number: 3352).