SUPPLEMENTARY INFORMATION

Protein-protein interactions of Human Glyoxalase II: findings of a reliable docking protocol

Roberta Galeazzi*[1], Emiliano Laudadio[1], Emanuele Falconi[1], Luca Massaccesi[1], Luisa Ercolani[2#], Giovanna Mobbili[1], Cristina Minnelli[1], Andrea Scirè[1], Laura Cianfruglia[2], Tatiana Armeni[2]

Assessment of Protein-Protein docking protocol

The reliability of the used protein-protein docking protocol was firstly assessed starting from [64-67]. We decided to start from three docking servers that are the most suitable and accurate ones to predict the blind protein-protein association: the first server which we tested is *ZDOCK*, which considers a combination of shape complementarity, Coulomb electrostatic term and free desolvation energy. This webserver uses the FFT (Fast Fourier Transformation) algorithm, which selects the best generated protein-protein complexes by selecting those with favorable desolvation and electrostatic energies.

The second server is *ClusPro 2.0*, which performs a rigid docking. It makes a quick estimate of potentials, such as the Atomic Contact Potential and the electrostatic energies for filtering, a ranking based on the clustering properties of low free energy complexes; furthermore it makes a short side chain minimization using CHARMM to remove the clash on the docking interface.

The third protein-protein server tested is the *Patch dock server* (bioinfo3d.cs.tau.ac.il/PatchDock/), which is a geometry-based molecular docking algorithm aimed at finding docking transformations that yield good molecular shape complementarity. Such transformations, when applied, induce both wide interface areas and small amounts of steric clashes. A wide interface is ensured to include several matched local features of the docked molecules that have complementary characteristics using the Connolly dot surface representation and classifying them as into concave, convex and flat patches. Each candidate patch is further evaluated by a scoring function that considers both geometric fit and atomic desolvation energy [68-69]. Finally, RMSD (root mean square deviation) clustering is applied to the candidate solutions to discard redundant solutions. The results obtained from this server were further refined with the subsequent submission to the associate server FireDock, which provide a further refinement of both the score function and of the complexes geometries. These refinements include both conformational optimizations of the later chains and of the rigid body orientation. A

further improvement has been obtained optimizing the geometry and the energy of the docked predicted complexes with AMBER force field [70-71].

In the first part of this work, we predicted protein-protein association, starting from these three most used servers in this field. In order to evaluate the used protocol, as test proteins, we choose *Starter proteins* and their corresponding crystallographic structures from CAPRI experiments [http://www.ebi.ac.uk/msd-srv/capri]. More precisely, the following couples were considered: Cellulosomal Scaffolding Protein *A and Endo-1,4-Beta-xylanase Y* (4BXG), *g-type Lysozyme* and its *Inhibitor (4G9S); Colicin D* e *Colicin D immunity protein (1V74)* and *Interferon-induced-guanylate-binding protein 1(2B92)*. In the last case, we generated the starters from the corresponding final crystallographic complex. In all cases, the proteins were freely roto-translated in the xyz space and crystallization water was removed before proceeding with the macromolecular docking.

Z-DOCK results (zdock.umassmed.edu): None of the four macromolecular complexes was predicted among the first ten Z-DOCK previsions (highest scored ones). Then, we searched among the 2000 predicted docked structures performing first a Cluster analysis and then performing a single point energy calculation of all the representative structures of each cluster using a state-of art force field AMBER within the Macromodel framework [72]. All the refined structures were then ranked according to their increasing energy and once again the lowest energy complexes were still different from the crystallographic ones.

ClusPro results (cluspro.bu.edu/login.php) [73]: represents the first fully automated, web-based program for the computational docking of protein structures. Its docking algorithms evaluate billions of putative complexes, retaining a preset number with favorable surface complementarities. A filtering method is then applied to this set of structures, selecting those with good electrostatic and desolvation free energies for further clustering. The program output is a short list of putative complexes ranked according to their clustering properties. Particularly the results are organized in four blocks in which the representative structures of each cluster are reported together with its energetic value. In the first block the best previsions are reported evaluated according their hydrophobicity; in the second block, the best previsions according their electrostatic energy are classified; in the third the best previsions according their van der Waals energy contribution are reported; finally in the fourth, the most probable complexes considering all the three energetic contribution are put. However, despite to the promising results we must discard it for our subsequent study of GlxII association since it cannot consider metals in its calculations and we need to keep the two zinc atoms of the GlxII binding site.

Patch dock results (bioinfo3d.cs.tau.ac.il/PatchDock/) [74]: using this geometry-based molecular docking algorithm, we evaluated the complexes scoring function that considers both geometric fit and atomic desolvation energy. Then, RMSD (root mean square deviation) clustering is applied to the candidate solutions to discard redundant solutions. The results obtained from this server were the most accurate, and we further improve them with the subsequent submission to the associate server FireDock, which provides a refinement of the score function and of the complexes geometries. These refinements include both conformational optimizations of the later chains and of the rigid body orientation. For the evaluation of the results obtained from these three servers, we carried out four rounds, one for each couple using all the three servers described above. Each predicted complex has been superimposed with the corresponding crystallographic one. In all cases the best results were obtained using the combination PatchDock/FireDock server since in this case the best prevision gained the first positions in the top ten server previsions (**Figure S1 and S2).** A further improvement has been obtained optimizing the geometry and the energy of the docked predicted complexes with AMBER force field [71].

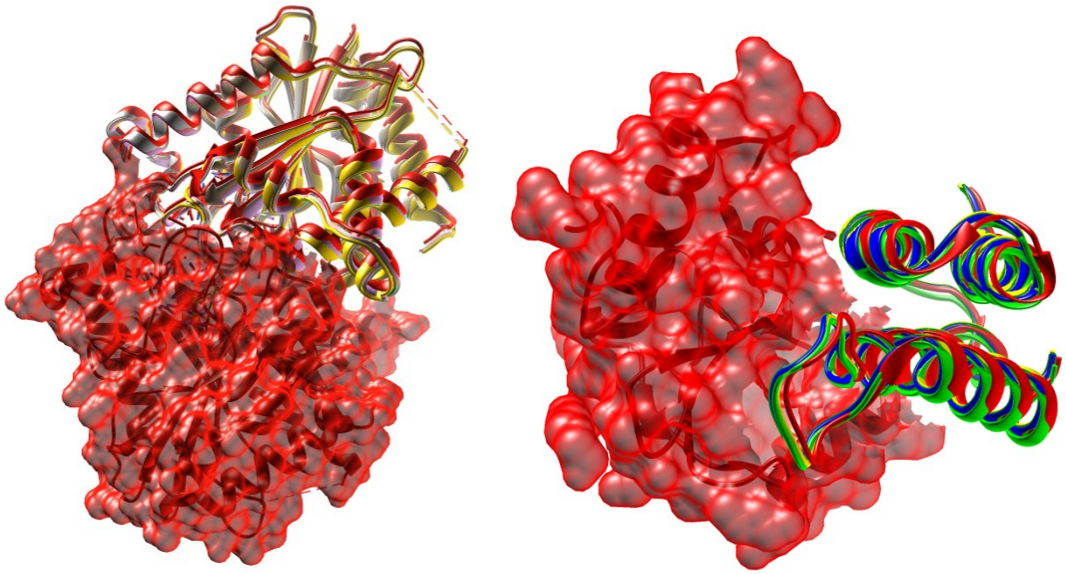

Figure S1. RMSD superimposition of the best Protein-Protein docking previsions obtained from PatchDock/FireDock with the crystallographic structure of *(a) Interferon-induced guanylate-binding protein 1; (b) Colicin D and its immunity protein;* the crystallographic reference structure is represented in red.

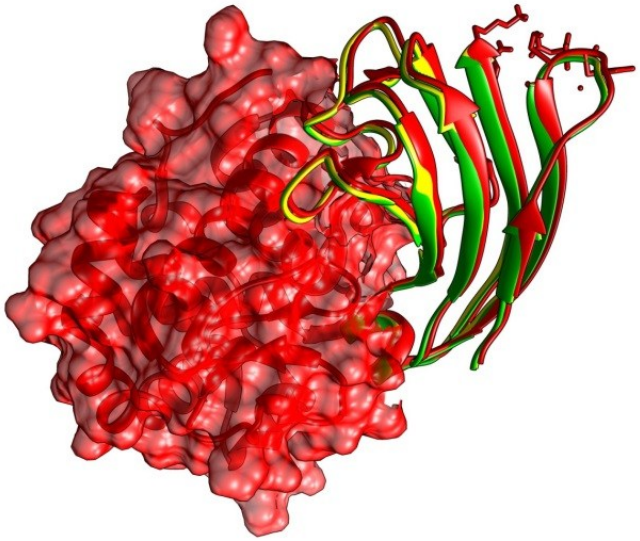

Figure S2. RMSD superimposition of the best Protein-Protein docking previsions obtained from PatchDock/FireDock with the crystallographic structure of g-type lysozyme and its inhibitor; the crystallographic reference structure is represented in red.

Analyzing the results obtained, we proceeded to carry out the Protein-Protein docking for GlxII using the PatchDock/FireDock approach followed by AMBER energy refinement and MD stabilization (see Methods section in the main paper for more details).

Additional PatchDock/FireDock results for GlxII-actin complexes:

in presence of GSH: Among all the structures (over 1000 reported from PatchDock), the greatest variability in the score are the top 20, the same that we refined by FireDock server. In this group of structures, we can observe a score from 15896 to 13576. After this limit, the score value varies only for few tens, thus suggesting that the focus for the refinement must be within this group. After Firedock refinement, the global energy predicted from the score function varies from -23.59 Kcal/mol (number 1) to 1.36 kcal/mol (number 20).

in absence of GSH: Among all the structures (over 1000 reported from PatchDock), the greatest variability in the score are the top 20, the same that we refined by FireDock server. In this group of structures, we can observe a score from 15738 to 13272. After this limit, the score value varies only for few tens, thus suggesting that the focus for the refinement must be within this group. After Firedock refinement, the global energy predicted from the score function varies from -12.43 Kcal/mol (number 1) to 34.85 kcal/mol (number 20).

GAPDH experimental data (never published before)

Human Glyoxalase II does not promote S-glutathionylation of GAPDH, as it is clear in Figure S3.

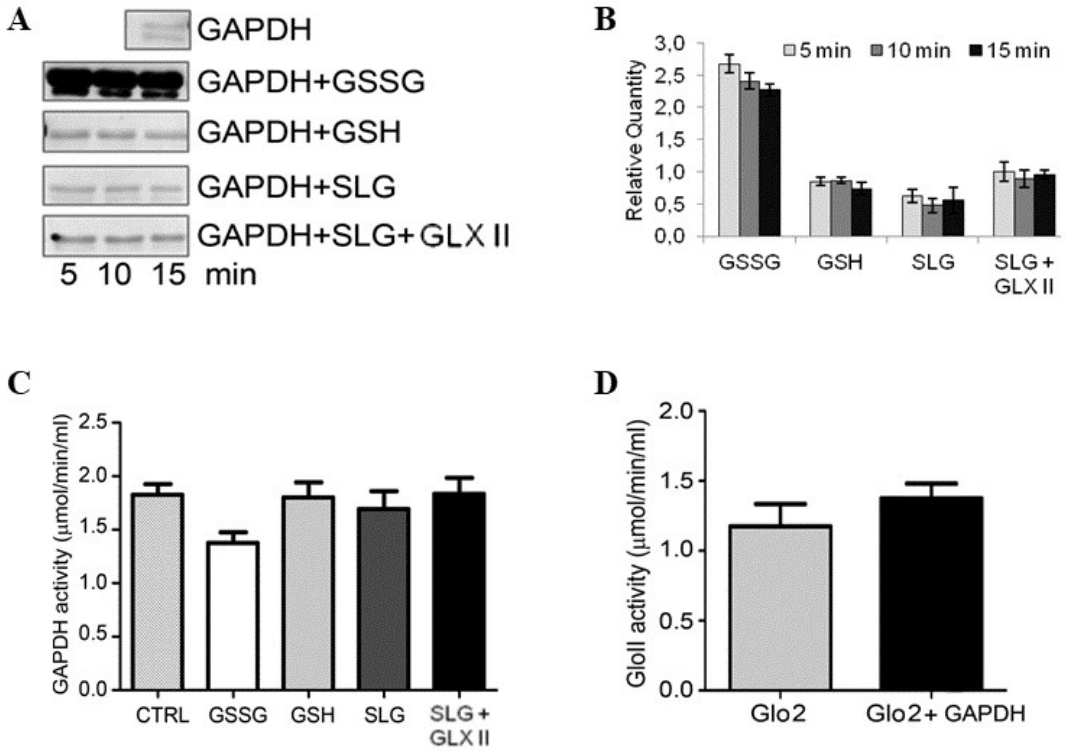

Figure S3. A. Immunoblot with anti-GSH antibody of 20 µg GAPDH protein incubated at 25°C with 1 mM GSSG, 1 mM GSH, 1 mM SLG or 0.2 U GlxII plus 1 mM SLG for 5-10-15 minutes. **B**. Densitometry measurements of immunoblot reported in A. Results are mean values ± s.d. from at least n=4 independent experiments. Statistical analyses were performed using Student's t-test comparing proteins incubated with SLG alone and with SLG + GlxII. **C.** Enzymatic activity of GAPDH incubated with 1 mM GSSG, 1 mM GSH, 1 mM SLG or 0.2 U GlxII plus 1 mM SLG at 25°C. Activity is given as µmol/min/ml. Results reported are the mean ± s.d. at least n=4 different experiments. Statistical analyses were performed using Student's t-test comparing GAPDH activity when incubated with the described molecules with respect to the control. **D.** GlxII activity in presence or absence of GAPDH protein. Statistical analyses were performed using Student's t-test compared GlxII activity alone with respect GlxII plus MDH

Table S1. MM/PBSA energy for GlxII-GADPH complexes in presence and in absence of GSH

ΔG binding for GlxII-GADPH (GSH)	-45.02 ± 7.95
ΔG binding for GlxII-GADPH (no GSH)	-34.76 ± 9.50

BLAST Alignment of human and rabbit Actin

```
>sp|P68135|ACTS_RABIT Actin, alpha skeletal muscle OS=Oryctolagus cuniculus
MCDEDETTALVCDNGSGLVKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEA
QSKRGILTLKYPIEHGIITNWDDMEKIWHHTFYNELRVAPEEHPTLLTEAPLNPKANREK
MTQIMFETFNVPAMYVAIQAVLSLYASGRTTGIVLDSGDGVTHNVPIYEGYALPHAIMRL
DLAGRDLTDYLMKILTERGYSFVTTAEREIVRDIKEKLCYVALDFENEMATAASSSSLEK
SYELPDGQVITIGNERFRCPETLFQPSFIGMESAGIHETTYNSIMKCDIDIRKDLYANNV
MSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWIT
KQEYDEAGPSIVHRKCF
```

```
>sp|P68133|ACTS_HUMAN Actin, alpha skeletal muscle OS=Homo sapiens
MCDEDETTALVCDNGSGLVKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEA
QSKRGILTLKYPIEHGIITNWDDMEKIWHHTFYNELRVAPEEHPTLLTEAPLNPKANREK
MTQIMFETFNVPAMYVAIQAVLSLYASGRTTGIVLDSGDGVTHNVPIYEGYALPHAIMRL
DLAGRDLTDYLMKILTERGYSFVTTAEREIVRDIKEKLCYVALDFENEMATAASSSSLEK
SYELPDGQVITIGNERFRCPETLFQPSFIGMESAGIHETTYNSIMKCDIDIRKDLYANNV
MSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWIT
KQEYDEAGPSIVHRKCF
```

Alignment statistics for match #1

Score	Expect	Method	Identities	Positives	Gaps
790 bits(2039)	0.0	Compositional matrix adjust.	377/377(100%)	377/377(100%)	0/377(0%)

```
Query  1    MCDEDETTALVCDNGSGLVKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEA  60
            MCDEDETTALVCDNGSGLVKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEA
Sbjct  1    MCDEDETTALVCDNGSGLVKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEA  60

Query  61   QSKRGILTLKYPIEHGIITNWDDMEKIWHHTFYNELRVAPEEHPTLLTEAPLNPKANREK  120
            QSKRGILTLKYPIEHGIITNWDDMEKIWHHTFYNELRVAPEEHPTLLTEAPLNPKANREK
Sbjct  61   QSKRGILTLKYPIEHGIITNWDDMEKIWHHTFYNELRVAPEEHPTLLTEAPLNPKANREK  120

Query  121  MTQIMFETFNVPAMYVAIQAVLSLYASGRTTGIVLDSGDGVTHNVPIYEGYALPHAIMRL  180
            MTQIMFETFNVPAMYVAIQAVLSLYASGRTTGIVLDSGDGVTHNVPIYEGYALPHAIMRL
Sbjct  121  MTQIMFETFNVPAMYVAIQAVLSLYASGRTTGIVLDSGDGVTHNVPIYEGYALPHAIMRL  180

Query  181  DLAGRDLTDYLMKILTERGYSFVTTAEREIVRDIKEKLCYVALDFENEMATAASSSSLEK  240
            DLAGRDLTDYLMKILTERGYSFVTTAEREIVRDIKEKLCYVALDFENEMATAASSSSLEK
Sbjct  181  DLAGRDLTDYLMKILTERGYSFVTTAEREIVRDIKEKLCYVALDFENEMATAASSSSLEK  240

Query  241  SYELPDGQVITIGNERFRCPETLFQPSFIGMESAGIHETTYNSIMKCDIDIRKDLYANNV  300
            SYELPDGQVITIGNERFRCPETLFQPSFIGMESAGIHETTYNSIMKCDIDIRKDLYANNV
Sbjct  241  SYELPDGQVITIGNERFRCPETLFQPSFIGMESAGIHETTYNSIMKCDIDIRKDLYANNV  300

Query  301  MSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWIT  360
            MSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWIT
Sbjct  301  MSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWIT  360

Query  361  KQEYDEAGPSIVHRKCF  377
            KQEYDEAGPSIVHRKCF
Sbjct  361  KQEYDEAGPSIVHRKCF  377
```

References (numbering is progressive from main paper)

64. Chen R, Li L, and Weng Z. ZDOCK: an initial-stage protein-docking algorithm. *Proteins* 52*:*80-87, 2003.

65. Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJ, Vajda S, Vakser I, and Wodak SJ. CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins* 52*:* 2-9, 2003.

66. Wiehe K, Pierce B, Mintseris J, Tong WW, Anderson R, Chen R, and Weng Z. ZDOCK and RDOCK performance in CAPRI rounds 3, 4, and 5. *Proteins* 60*:* 207-213, 2005.

67.Mintseris J, Pierce B, Wiehe K, Anderson R, Chen R, and Weng Z. Integrating statistical pair potentials into protein complex prediction. *Proteins* 69*:* 511-520, 2007.

68**.** Duhovny D, Nussinov R, Wolfson HJ. Efficient Unbound Docking of Rigid Molecules. In Gusfield et al., Ed. Proceedings of the 2'nd Workshop on Algorithms in Bioinformatics(WABI) Rome, Italy, Lecture Notes in Computer Science 2452, pp. 185-200, Springer Verlag, 2002.

69. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucl.Acids. Res. 33: W363-367, 2005.
70. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995). "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules". *J. Am. Chem. Soc.* **117**: 5179–5197].
71.Case DA, Darden TA, Cheatham TE, III, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts B, Hayik S, Roitberg A, Seabra G, Swails J, Goetz AW, Kolossvai I, Wong KF, Paesani F, Vanicek J, Wolf RM, Liu J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh M-J, Cui G, Roe DR, Mathews DH, Seetin MG, Salomon-Ferrer R, Sagui C, Babin V, Luchko T, Gusarov S, Kovalenko A, and Kollman PA. AMBER 11. *University of California, San Francisco; ambermd.org* 2011; AMBER Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules *J. Am. Chem. Soc.* **117**: 5179–5197.

72. Macromodel framework, Shroedinger Suite 2015

73. ClusPro: Kozakov D, Beglov D, Bohnuud T, Mottarella S, Xia B, Hall DR, Vajda, S. How good is automated protein docking? *Proteins: Structure, Function, and Bioinformatics*, 2013 Aug; Kozakov D, Brenke R, Comeau SR, Vajda S. PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins*. 2006 Aug 24; Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*. 2004 Jan 1;
74. PatchDock: Duhovny D, Nussinov R, Wolfson HJ. Efficient Unbound Docking of Rigid Molecules. In Gusfield et al., Ed. Proceedings of the 2'nd Workshop on Algorithms in Bioinformatics(WABI) Rome, Italy, Lecture Notes in Computer Science 2452, pp. 185-200, Springer Verlag, 2002; Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucl.Acids. Res. 33: W363-367, 2005; Zhang C., Vasmatzis G., Cornette J.L., DeLisi C. Determination of atomic desolvation energies from the structures of crystallized proteins. J. Mol. Biol. 1997, **267**,707-26.