Supplemental Data

Supplemental Materials and Methods

Yeast 2-hybrid assay

All the fragments and full-length CRP and GBP cDNAs (without their signal sequences) were each fused to the DNA-binding domain of Gal4 in the bait plasmid pGBKT7 (BD Biosciences), or to the activation domain of Gal4 in the prey plasmid pGADT7-Rec (BD Biosciences) and cotransformed into *S. cerevisiae*, strain AH109. For selection, synthetic complete (SC) media lacking Leu and Trp (SC-Trp-Leu) or lacking Leu, Trp, His and adenine (QDO medium) were used. Transformants containing bait and prey plasmids were selected on SC-Trp-Leu by incubation for 3.5 days at 30 °C. Resulting colonies were suspended in water and replated on SC-Trp-Leu and QDO agar at 30 °C for up to a maximum of 7 days. The negative control was co-transformed with a recombinant plasmid and an empty prey or bait plasmid. The positive control was co-transformed with a plasmid expressing the full-length Gal4 transcriptional activator together with the empty pGADT7-Rec vector. The prey and bait plasmids were reciprocally cloned into pGADT7 and pGBKT7 vectors and re-cotransformed to verify the interactions.

Surface plasmon resonance analysis (SPR)

The binding between GBP and its ligands (GlcNAc, LPS, ReLPS and lipid A from *Salmonella minnesota*, R595; List Biologicals, UK) were characterized by real-time biointeraction analyses using a Biacore 2000 instrument (Biacore AB). For GBP-GlcNAc interaction, GlcNAc-BSA (Dextra Labs, UK) was diluted to 10 µg/ml with 10 mM sodium acetate, pH 4.0, and immobilized on the surface of a CM5 sensor chip (Biacore AB) using amine-coupling chemistry according to the manufacturer's specifications. Binding of GBP to immobilized GlcNAc-BSA was measured at a flow rate of 20 µl/min in TBS. Regeneration of the surfaces was achieved by injecting 20 µl of the running buffer containing 300 mM GlcNAc. For GBP-LPS interaction, LPS, ReLPS and lipid A were diluted to 0.25 mg/ml in 20 mM sodium phosphate, 150 mM NaCl, pH 7.4 and immobilized on the surface of an HPA sensor chip according to the manufacturer's specifications. Binding of GBP to the immobilized ligands was measured at a flow rate of 20 µl/min in TBS (10 mM Tris, 150 mM NaCl, pH 7.4). Regeneration of the chip surface was achieved by injection of 20 μ l 0.1 M NaOH. The binding affinities were calculated using BiaEvaluation software, version 4.1.

ELISA to test for bacterial ligand-binding

The GBP ligands (LPS, ReLPS, lipid A, LTA and GlcNAc-BSA) were incubated overnight in binding buffer (PBS, 3.2 mM Na2HPO4, 0.5 mM KH2PO4, 1.3 mM KCl, 135 mM NaCl, pH 7.4.) on 96-well PolysorpTM (Nunc) microplates. After washing off excess ligands, the unbound sites were blocked with 1% BSA and incubated at 25 °C for 2 h. Serially diluted GBP samples (with or without pre-incubation with GlcNAc) were then added to each well and incubated at 25 ◦ C for 2 h. The rabbit antiserum against GBP was added after washing each well. Subsequently, horseradish peroxidase-linked anti-rabbit IgG antibody was added and incubated for 1 h. Peroxidase enzyme activity was determined after adding ABTS (Amersham, GE Healthcare) as substrate and measured at 405 nm.

Hydrogen deuterium exchange mass spectrometry (HDMS)

Mixtures containing GBP+GlcNAc, GBP+lipid A, GBP+CRP, GBP+CRP+lipid A, and GBP alone (as control), were deuterated by diluting 2 μl of protein solution (10 mg/ml, 0.4 mM) in TBS containing 10 mM Tris, 150 mM sodium chloride, pH 7.4 to 18 μ l D₂O (Sigma). Solutions were then quenched after timed intervals of 30 s, 1, 2, 5 and 10 min with 180 μl of 0.1% trifluoroacetic acid, pH 2.5 (TFA, Sigma). Quenched reactions were digested for 5 min with immobilized pepsin (Thermo Scientific). Solutions were spun down for 2 min at 8000g and 20 μl aliquots of digested solutions were flash-frozen in liquid nitrogen and kept at -80 °C until analysis by matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI TOF MS). Undeuterated samples were included as a negative control. The undeuterated sample was similarly treated (replacing D_2O with TBS) to obtain the sequence of all the pepsin digested fragments by MS/MS sequencing on the MALDI TOF/TOF 4800 instrument (ABI Biosystems). One sample was allowed to exchange with deuterated buffer for 24 h to allow for complete deuteration of solvent exposed regions of the protein. This was used to determine the amount of back-exchange in our experiments.

For MALDI TOF MS analysis, the reaction aliquots were thawed and 0.5 μl was mixed with 0.5 μl of mass spectrometry matrix solution (15 mg/ml α-cyano-4-hydroxycinnamic acid in 1:1:1 ethanol: acetonitrile: 0.1% TFA, pH 2.5). 0.5 μl of this mixture was spotted on the MALDI plate and quickly dried under vacuum and analysed on the MALDI-TOF mass spectrometer. Spectra were calibrated using Data Explorer (Applied Biosystems) with internal peptide masses 974.51 and 1492.72 and the centroid of the peptide envelopes were measured using Decapp Mass Spec Isotope Analyzer (UCSD, Jeffrey G. Mandell and Elizabeth A. Komives) (1). Back exchange was found to be $\sim 63\%$, so all centroid values were multiplied by a back exchange factor of 2.67 to calculate the experimental deuterium exchange levels.

A total of 59 pepsin-digested peptide fragments for GBP were generated. These were sequenced by MS/MS sequencing and were found to encompass 92% of the GBP amino acid sequence. Out of these peptides, 20 covering 73% with good signal-to-noise ratio under HDMS experimental conditions were chosen for further analysis, and the extent of deuterium exchange was plotted against exchange time for GBP, GBP+GlcNAc, GBP+lipid A, GBP+CRP and GBP+CRP+lipid A. A total of 70 pepsin-digested peptide fragments for CRP were generated. These were sequenced by MS/MS sequencing and were found to encompass 98% of the CRP amino acid sequence. Out of these peptides, 19 covering 89% with good signal-to-noise ratio under HDMS experimental conditions were chosen for further analysis, and the extent of deuterium exchange was plotted against exchange time for CRP, GBP+CRP and GBP+CRP+lipid A.

Molecular modeling and docking

The Molecular Operating Environment (MOE) software (2) was used to build the GBP model. The sequences of GBP and TL-1 were aligned using the Blosum62 substitution matrix. The stochastic model building algorithm of MOE uses an adapted Boltzmann-weighted randomized modeling procedure (3) combined with specialized logic for the proper handling of insertions and deletions (4). The side chain data was assembled from an extensive rotamer library generated by systematic clustering of high-resolution PDB data. MOE created a collection of 10 independent intermediate homology models, which contain different loop candidates and side chain rotamers. Loops were modeled first, in random order. For each loop, a contact energy function was used to select the best candidate. The intermediate models were scored by the electrostatic solvation energy scoring function, calculated via Generalized Born/Volume Integral (GB/VI) methodology. Using molecular mechanics and AMBER99 force field (5) with electrostatic reaction field correction, the best scoring model was subjected to energy minimization (6). The final model was inspected using MOE's Protein Geometry stereochemical quality evaluation tools to confirm that the model's stereochemistry is reasonably consistent with typical values found in crystal structures. The Ramachandran plot (generated by RAMPAGE (6)) shows that the outlier residues listed remain close to the boundaries of the permitted Psi-Phi values, which are indicated by the light blue contours.

Saccharides and lipid A molecules were constructed using the Biopolymer module in Insight II (7) and docked in MOE (2). MOE has been previously successfully applied in the modeling of saccharide-receptor interactions (8-10). Initial fitting of the ligands (flexible ligand and rigid receptor docking) was performed using MOE Site Finder and Docking modules (2). The top 30 semi-flexible docking poses generated by MOE were rank-ordered, and further energy-refined using the Docking module of Insight-II and AMBER99 force field (11). Receptor residues within a 5 Å radius of the docked ligand were considered as flexible during the structural refinement and relaxation. During the refinement, the lowest energy combination of side chain rotamers of the flexible residues was determined in several iterative cycles followed by careful minimization of the energy of the resulting ligand-receptor complex. A non-bonding interaction distance cutoff set to 15 Å was employed and dielectric constant of 2 was used to take into account dielectric shielding effects in proteins. Minimization of the complex was carried out by relaxing the structure gradually, starting with hydrogen atoms, then proceeding with all ligand atoms, followed by receptor residue side chains and concluding the relaxation with freeing of all atoms including the protein backbone. In the geometry optimization, a sufficient number of conjugate gradient iterative cycles, with a convergence criterion set to the average gradient of 0.01 kcal x mol⁻¹ x \AA ⁻¹, was used. The ligand-receptor interaction energy E_{int} was computed as the sum of electrostatic and van der Waals terms as defined in the AMBER99 force field (11).

Protein-protein docking

400 rigid body docking structures were generated and 40 highest scoring structures were refined by semi-flexible simulated annealing involving active and passive residues. Passive residues surround the active ones and are located in a layer defined by 3 Å interatomic distance from the active residues. The total energy calculated by molecular mechanics of the dimer (using a force field of CNS program (12,13)), protein-protein interaction energy and protein desolvation energy were considered as a weighted sum of parameters in an empirical scoring function that was used for the rank-ordering of the generated GBP-CRP dimers.

The scoring function for the docked structure is calculated as follows:

$$
Score = a \cdot E_{tot} + b \cdot E_{int} + c \cdot E_{desolv}
$$

where $a = -0.70$, $b = -0.15$, $c = -0.15$. E_{tot} is total molecular mechanics energy in the CNS force field, E_{int} is interaction energy between monomers in the heterodimer, E_{desolv} is desolvation energy change of monomers due to the heterodimer formation

Pyrogene assay for LPS endotoxicity

To test the potential anti-LPS effects of GBP and CRP, the endotoxicity of LPS with and without these proteins were measured using the PyroGene kit (Lonza Inc.). This kit measures the endotoxicity of LPS by using a recombinant enzyme, Factor C (rFC) from the horseshoe crab. Upon encountering LPS, the rFC is activated and it hydrolyses a substrate to produce a fluorescent product, thereby reporting on the endotoxin activity from as low as 0.05 EU/ml of LPS. A total volume of 100 μl, constituted by 50 μl of LPS at different EU (endotoxin units), 25 μl of GBP/CRP ranging from 0.01 to 10 μM and sterile water were added into fluorescence microplate wells (Nunc). The reactions were pre-incubated at 37 °C for 10 min. The working reagent was prepared by mixing rFC solution, the assay buffer and fluorogenic substrate at a ratio of 1:4:5, respectively. Then, 100 μl of the working solution was added to the test mixture just before the fluorescence measurement. The reaction was further incubated for 1 h and the fluorescence was measured again. The EU of the test mixture was compared against a standard curve of 0.01-10 EU of LPS.

References

- 1. Mandell, J. G., Falick, A. M., and Komives, E. A. (1998) *Anal Chem* **70**, 3987-3995
- 2. Molecular Operating Environment version 2007.12. The Chemical Computing Group Inc., Montreal
- 3. Levitt, M. (1992) *J Mol Biol* **226**, 507-533
- 4. Petsko, B. G. A., and Ringe, D. (2004) Structure from sequence : Homology modelling. in *Protein Structure and Function*, 2 Ed., New Science Press
- 5. Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) *J Comput Chem* **26**, 1668-1688
- 6. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) *Proteins* **50**, 437-450
- 7. Low, D. H. P., Ang, Z. W., Yuan, Q., Frecer, V., Ho, B., Chen, J., and Ding, J. L. (2009) *PLoS ONE* **4**, e6260
- 8. Daiyasu, H., Saino, H., Tomoto, H., Mizutani, M., Sakata, K., and Toh, H. (2008) *J Biochem* **144**, 467-475
- 9. Naberezhnykh, G. A., Gorbach, V. I., Likhatskaya, G. N., Davidova, V. N., and Solov'eva, T. F. (2008) *Biochemistry (Mosc)* **73**, 432-441
- 10. Vilar, S., Cozza, G., and Moro, S. (2008) *Curr Top Med Chem* **8**, 1555-1572
- 11. Wang JM, Cieplak P, and PA, K. (2000) *Journal of Computational Chemistry* **21**, 1049- 1074
- 12. Nilges, M. (1995) *J Mol Biol* **245**, 645-660
- 13. Nilges, M. (1996) *Curr Opin Struct Biol* **6**, 617-623

Ramachandran plot for the structure of CRP. The outlier residues listed remain close to the boundaries of the permitted Psi-Phi values, which are highlighted by the light blue contours, indicating that the structure has been reliably modeled. Black, orange and red boxes represent residues in favoured, allowed and outlier regions respectively (refer to table below).

List of Phi-Psi outliers in the CRP model.

Number of <u>residues</u> in favoured region (~98.0% expected) : 148 (68.5%) Number of residues in allowed region $(-2.0\%$ expected) : 55 (25.5%) Number of residues in outlier region : 13 (6.0%)

Horseshoe crab plasma was applied to an affinity column using Sepharose CL-6B as a matrix. The column was equilibrated with 10 mM Tris.Cl pH 8.8, 150 mM NaCl and extensively washed with the same buffer. GBP was eluted with the same buffer containing 0.4 M GlcNAc. Fractions 112-120 were pooled and used for further studies. Proteins in the crude extract, flowthrough (FT), washes and representative eluted fractions were resolved by 12% SDS-PAGE and stained with Coomassie Brilliant Blue, and immunoblotted (IB) with GBP antibody.

The chemical structure of lipoteichoic acid (LTA) **(A)** shows the presence of GlcNAc, a common target of GBP. ELISA shows that GBP binds LTA, and the binding is inhibited by the addition of GlcNAc **(B)**. "+GlcNAc" indicates that GBP has been preincubated with GlcNAc prior to binding analysis with LTA.

Supplemental Figure 4

GBP & CRP surfaces showing changes through GBP:ligand, GBP:CRP or CRP:GBP interaction. The amount of deuterium ions incorporated over time in the GBP and CRP peptide fragments indicates the propensity to interact (decrease) or undergo conformational changes (increase). 9 peptide fragments in GBP show significant differences in deuterium exchange when it was interacted with lipid A (2 fragments), GlcNAc (3 fragments) and CRP (4 fragments). Three peptide fragments in CRP show significant differences in deuterium exchange when it was interacted with GBP (see also Supplementary Table 1&2). Color of the arrows correspond to the respective surfaces in Figure 5D,E,F.

 a Score = a·E_{tot} + b·E_{int} + c·E_{desolv}, a = -0.70, b = -0.15, c = -0.15, the higher is the value of the score (more positive, less negative) the better is the dimer structure. $\mathsf{E}_{\scriptscriptstyle{\text{tot}}}$ is total energy calculated by molecular mechanics of the heterodimer in CNS force field, E_{int} is interaction energy between monomers in the heterodimer, $\mathsf{E}_{\mathsf{desolv}}$ is desolvation energy change of monomers due to the heterodimer formation, $\mathsf{S}_{\mathsf{bur}}$ is area of molecular
surface buried during the heterodimer formation.

Guided docking of the GBP-CRP (blue, red respectively) interaction based on HDMS-defined active residues. Blind/random docking utilizing all solvent-accessible active residues in the interaction was used for comparison. The accompanying table lists scores obtained for the best models from the guided and blind runs.

The interaction between GBP and LPS is independent of Ca²⁺. SPR showed that GBP's binding affinity to LPS did not change in (i,ii) the presence or (iii) absence of Ca²⁺.

Protein-protein interaction between PRRs, observed via yeast 2-hybrid

Weak interaction Very weak interaction

The PRR-interactome involves three core PRRs already known to form a complex: GBP, CRP and CL5-C (Ng et al, 2007). Serine proteases (CrC2/Bf, involved the complement activation and CrFC, which activates the coagulation cascade) have been shown to boost this PRR interactome formation (Le Saux et al, 2008). CL5C, a carcinolectin, is a homolog of the human ficolin and both proteins are involved in complement activation. HMC, hemocyanin (the oxygen-transport protein in plasma), coagulogen (the precursor of coagulin, the final product of coagulation) and PPOA (prophenol oxidase A) are also involved in the interactome, abridged by GBP.

GBP and/or CRP dissociate LPS micelles and unmask the endotoxicity of LPS. Increasing doses of LPS from **(A)** 0.5 EU/ml, **(B)** 1EU/ml, and **(C)** 2.0 EU/ml) were reacted with various concentrations of GBP or CRP or GBP+CRP, and the resulting endotoxic activities of LPS were assayed using the PyroGene kit. The activities were plotted and compared with the corresponding LPS concentrations (black dashed line in each graph). **(D)** The endotoxicity is significantly increased as GBP/CRP/GBP+CRP causes the relative increase in the unmasking of LPS micelles. A p-value < 0.05 as calculated by Student's *t* test was regarded statistically significant.

Supplemental Figure 9

Selected hTectonin library screened partners for interaction confirmation

ND denotes no significant difference in standard error of mean

ND denotes no significant difference in standard error of mean

Accession	Name	Function	Tissue
Q9H6Q3	Src-like-adaptor 2 isoform a	Adapter protein, which negatively regulates T-cell receptor (TCR) signaling. Inhibits T-cell antigen-receptor induced activation of nuclear factor of activated T-cells. May act by linking signaling proteins such as ZAP70 with CBL, leading to a CBL dependent degradation of signaling proteins.	Thymus, Hepatoma, Prostate
P43490	Nicotinamide phosphoribosyltransferase (Pre-B-cell colony- enhancing factor 1)	Catalyzes the condensation of nicotinamide with 5- phosphoribosyl-1-pyrophosphate to yield nicotinamide mononucleotide, an intermediate in the biosynthesis of NAD. It is the rate limiting component in the mammalian NAD biosynthesis pathway	Expressed in large amounts in bone marrow, liver tissue, and muscle. Also present in heart, placenta, lung, and kidney tissues.
P14598	Neutrophil cytosol factor 1	NCF2, NCF1, and a membrane bound cytochrome b558 are required for activation of the latent NADPH oxidase (necessary for superoxide production). Defects in NCF1 are the cause of chronic granulomatous disease autosomal recessive cytochrome-b-positive type 1 (CGD1) [MIM:233700]. Chronic granulomatous disease is a genetically heterogeneous disorder characterized by the inability of neutrophils and phagocytes to kill microbes that they have ingested. Patients suffer from life- threatening bacterial/fungal infections.	Widely expressed.

Supplemental Table 3 Putative hTectonin interaction partners via yeast 2-hybrid library screen