Supporting Information

Kim et al. 10.1073/pnas.1106840108

SI Text

Gene Cloning and Protein Expression. *hetR* sequences from *Fischerella* sp. MV11, *Nostoc punctiforme* PCC 73102, *Anabaena* sp. strain PCC 7120 *and Trichodesmium erythraeum* IMS101 were used to design primers to amplify the HetR ORFs by polymerase chain reaction (PCR) using genomic DNAs as templates. The amplified product was directly cloned in the vector pMCSG19 (1) using a modified ligation-independent cloning protocol (2). The HetR R62E, K73E, and R74E point mutations were prepared using a two-step PCR method and overlapping mutant primers. All PCR-cloned genes were sequenced at the University of Chicago Cancer Research DNA Sequencing Facility.

pMCSG19 encodes the maltose binding protein (MBP) sequence followed in order by: (*i*) a TVMV protease cleavage site, (*ii*) a hexa-histidine tag (His_6-tag), (*iii*) a TEV protease cleavage site, and (*iv*) the target protein sequence. Upon cleavage with TEV protease, three residues (SerAsnAla) are added at the N terminus of the target protein. HetR proteins were produced as a fusion protein MBP-His_6-HetR in *Escherichia coli* BL21 (DE3) carrying plasmids pMAGIC, that encode three rare *E. coli* tRNAs (Arg [AGG], Arg[AGA], and Ile [ATA]), and pRK1037 (3) (Scientific Reagents, Inc.) that encodes the TVMV protease. Cells were grown with ampicillin and kanamycin at 100 and 30 μg/mL, respectively, to an optical density at 600 nm of 0.6 and isopropyl-β-D-thiogalactoside (IPTG, 0.4 mM) was added to induce expression of the fusion protein. MBP was removed in situ upon production of the protease encoded on plasmid pRK1037. This construct provided an N-terminal His_6-tag separated from the desired gene by a TEV protease recognition sequence. Native HetR protein was expressed in transformed BL21 pMAGIC cells grown in LB at 37 °C to A_{595} of 1.0, when protein expression was induced with 1 mM IPTG. The cells were then incubated at 18 °C overnight.

A SeMet derivative of the expressed protein was prepared as described earlier (4). The transformed BL21 pMAGIC cells were grown at 37 °C in M9 minimal medium supplemented with 0.4% (w/v) glucose, 8.5 mM NaCl, 0.1 mM $CaCl_2$, 2 mM $MgSO_4$, and 1% (w/v) thiamine. At an A_{595} of 1.0–1.5, 0.01% (w/v) leucine, isoleucine, lysine, phenylalanine, threonine, and valine were added to inhibit methionine synthesis and encourage SeMet incorporation. SeMet (90 mg) was added to one liter of culture and protein expression was induced with 1 mM IPTG. The cells were then incubated at 18 °C overnight. The harvested cells, containing either native or SeMet HetR, were resuspended in lysis buffer [500 mM NaCl, 5% (v/v) glycerol, 50 mM Hepes pH 8.0, 10 mM imidazole, 10 mM 2-mercaptoethanol] and stored at −80 °C.

Protein Purification. Both native and SeMet HetR proteins from *Fischerella* sp. MV11, *Nostoc punctiforme* PCC 73102, *Anabaena* sp. strain PCC 7120 and *Trichodesmium erythraeum* IMS101 were purified using the procedure described in Kim et. al. (5). The harvested cells were resuspended in lysis buffer supplemented with 1mg/mL lysozyme and 100 mL of protease inhibitor (Sigma, P8849) per 2 g of wet cells. This mixture was kept on ice for 20 min and then sonicated. The lysate was clarified by centrifugation at 36,000× *g* for 1 h and filtered through a 0.44 μm membrane. The clarified lysate was applied to a 5 mL HiTrap Ni-NTA column on an AKTAxpress system (GE Health Systems). The His_6-tagged protein was released with elution buffer (500 mM NaCl, 5% glycerol, 50 mM Hepes pH 8.0, 250 mM imidazole, 10 mM 2-mercaptoethanol) and the fusion tag was removed by treatment with recombinant His_6-tagged TEV protease (a gift from D. Waugh, NCI). Ni-NTA affinity chromatography was then used to remove the His_6-tag, uncut protein, and His_6-tagged TEV protease. The HetR protein was dialyzed against crystallization buffer containing 250 mM NaCl, 20 mM Hepes pH 8.0, 2 mM DTT and then concentrated to 20 mg/mL for crystallization using a Centricon with 5,000 MW cutoff (Millipore), flash-cooled and stored in liquid nitrogen.

DNA Binding Using EMSA. Recombinant HetR protein used for DNA binding was expressed as a fusion with a GST tag (using a pGEX 4T-1 vector). The protein was purified with a glutathione Sepharose 4 Fast Flow kit (from General Electric) and resuspended in PBS buffer (pH 7.4). The 29-bp self-complementary DNA oligonucleotide was obtained from Invitrogen, Inc. (5′-GTAGGCGAGGGGTCTAACCCCTCATTACC -3′). The DNA was used in gel shift reactions at a final concentration of 50 nM using a binding buffer containing 1 M NaCl, 100 mM Tris, 10 mM DTT, 2 mM $CaCl_2$ at pH 7.5 and HetR at 5 μM. The samples were loaded onto a 5% TBE gel and stained with SYBR Green (Roche, Inc.).

Size-Exclusion Chromatography. The molecular weight of native HetR protein in solution was determined by HPLC size-exclusion chromatography using a SRT SEC-150 (7.8 × 250 mm) column (Sepax Tech. Inc.) according to the method described previously (6). The column was calibrated with ribonuclease (13.5 kDa), conalbumin (75 kDa), aldolase (158 kDa), catalase (232 kDa), ferritin (440 kDa), and blue dextran (2,000 kDa). The chromatography was carried out at 22 °C at a flow rate of 1.2 mL/ min.

Crystallization. The HetR protein was crystallized using sitting drop vapor diffusion at 289 K in a CrystalQuick®96-well round-bottom plate (Greiner Bio-One North America, Inc.). A 400 nL droplet of protein (15 mg/mL) was mixed with a 200 nL droplet of crystallization reagent and allowed to equilibrate over 135 μL of crystallization reagent. The nanopipetting was performed using the Mosquito® nanoliter liquid handling system (TTP LabTech). The finished plate was then incubated at 16 °C within a RoboIncubator® automated plate storage system (Rigaku). Automated crystal visualization (Minstrel III, ® Rigaku) was utilized to locate several crystals, which were cryoprotected and flash cooled in liquid nitrogen. Although crystals were obtained for all HetR proteins produced in this study, X-ray diffraction quality crystals were found only for the HetR protein from *Fischerella* sp. MV11. The best crystals of SeMet-labeled protein were obtained at 20 mg/mL from 1 M $(NH_4)_2HPO_4$, 0.2 M NaCl, 0.1 M imidazole pH 8.0. These crystals belong to hexagonal space group $P6_5$ with unit cell dimension $a = 123.38$, $b = 123.38$, $c = 109.6$ Å and diffracted X-rays to better than 3.38 Å with a cryoprotectant of 28% sucrose. The best crystals of native HetR protein from *Fischerella* sp. MV11 were obtained at 15 mg/mL in 25% PEG4000, 50 mM Tris pH 7.5, 160 mM Na formate, 10 mM $MgCl_2$. The crystallization droplet also contained the self-complementary DNA oligonucleotide (TGGTGAGGGGTTAAACCCCTCACC), although the DNA was not detected in the crystal. The crystals of native protein belong to trigonal space group $P3_1$ with unit cell dimensions $a = 92.93$, $b = 92.93$, and $c = 97.65$ Å and diffracted X-rays to 3.0 Å in the presence of 10% glycerol as a cryoprotectant. Both crystal forms contained one HetR dimer in the asymmetric unit.

Data Collection. Diffraction data were collected at 100 K at the 19-ID beamline of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory (7). The single wavelength anomalous dispersion (SAD) data at 0.9793 Å (near the selenium absorption white line, 12.6605 keV) up to 3.38 Å were collected using inverse-beam geometry near the selenium absorption edge from a single ($0.2 \times 0.05 \times 0.05$ mm) SeMet labeled protein crystal. Before data collection, an X-ray fluorescence spectrum was recorded from each sample. The crystal was exposed for 5 sec. per 1.0° rotation of ω with the crystal to detector distance of 420 mm. The data were recorded on a CCD detector Q315r from Area Detector Scanning Corporation scanning a full 120° on ω. The higher resolution native data were also collected on the same detector, similarly at 0.9791 Å, with the crystal to detector distance of 450 mm, and 1.0° rotation per image frame of 3 sec exposure. The SBC-Collect program was used for data collection and visualization. Data collection strategy, integration, and scaling were performed with the HKL3000 program package (8). A summary of the crystallographic data can be found in Table 1.

Structure Determination. The 3.38 Å structure was determined by SAD phasing using HKL3000 (8). Se sites were determined by SHELXC (9) and SHELXD (9); the handedness was checked by SHELXE (9), phasing was done by MLPHARE (10), density modification was done by DM (11), and the first model was built by RESOLVE (12). The initial model from HKL3000 covered about half of the two chains in the asymmetric unit. Multiple cycles of manual model-building and adjustment using COOT (13) together with refinement using PHENIX (14) and REFMAC 5.5 (15) led to the final 3.38 Å model. The final model with R_{cryst} and R_{free} of 0.207 and 0.267, respectively, contained two HetR chains. The 3.00 Å native structure was determined by molecular replacement using MOLREP (16) on HKL3000. The search model was the B chain of the 3.38 Å structure. The structure was refined using PHENIX with twinning. After numerous iterations of manual adjustment and restrained refinement using Phenix, the model converged to contain two HetR molecules and five water molecules with R_{cryst} and R_{free} of 0.183 and 0.247. A total of 14 α-helices, 2 β-strands, 6 3_{10}-helices, 1 β-hairpin, and several connecting loops make up the HetR monomer (based on the B subunit structure). There was no interpretable electron density for N- and C-terminal residues as well as several loop regions and these parts of the protein are missing in the final model: Chain A: N-terminal SNA-1, 213–219, 279–285, 297-C-terminus; Chain B: N-terminal SNA-2, 147–149, 201–202, 214–217, 279–285, 297-C terminus. All residues fell within acceptable regions of a MolProbity Ramachandran plot, with 92.8% of all residues in favored regions. Refinement details are shown in Table 1.

Modeling RGSGR Peptide on HetR. The HetR structure was first analyzed to identify cavities with sufficient accessibility and volume to accommodate the pentapeptide RGSGR by comparing the HetR surface to existing protein-peptide binding surfaces in the PDB (17). Three conformations of the RGSGR peptide were modeled from existing pentapeptide structures that were bound to pockets shaped similarly to those on HetR, each representing the most plausible conformation for the three sites. The peptides were then posed in extended conformation into the binding site by performing rigid body rotations and translations to minimize steric clashing with HetR. Once a suitable pose was identified, a short conjugate gradient minimization procedure was performed (18). Solvent atoms were excluded during modeling.

1. Donnelly MI, et al. (2006) An expression vector tailored for large-scale, high-throughput purification of recombinant proteins. *Protein Expr Purif* 47:446–454.
2. Stols L, et al. (2002) A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expr Purif* 25:8–15.
3. Nallamsetty S, et al. (2004) Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro. *Protein Expr Purif* 38:108–115.
4. Walsh MA, Dementieva I, Evans G, Sanishvili R, Joachimiak A (1999) Taking MAD to the extreme: ultrafast protein structure determination. *Acta Crystallogr D BiolCrystallogr* 55:1168–1173.
5. Kim Y, et al. (2004) Automation of protein purification for structural genomics. *J Struct Funct Genomics* 5:111–118.
6. Nocek B, et al. (2005) Crystal structures of delta1-pyrroline-5-carboxylate reductase from human pathogens *Neisseria meningitides* and *Streptococcus pyogenes*. *J Mol Biol* 354:91–106.
7. Rosenbaum G, et al. (2006) The structural biology center 19ID undulator beamline: facility specifications and protein crystallographic results. *J Synchrotron Radiat* 13:30–45.
8. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M (2006) HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr* 62:859–866 (in eng).
9. Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* 58:1772–1779.
10. Otwinowski Z (1991) *Daresbury Study Weekend Proceedings*, (SERC Daresbury Laboratory, Warrington, United Kingdom), pp 80–85.
11. Cowtan K (1994) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* 31:34–38.
12. Terwilliger TC (2003) Improving macromolecular atomic models at moderate resolution by automated iterative model building, statistical density modification and refinement. *Acta Crystallogr D Biol Crystallogr* 59:1174–1182.
13. Emsley P, Cowtan K (2004) COOT: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
14. Adams PD, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213–221.
15. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240–255.
16. Vagin A, Teplyakov A (1997) MOLREP: an automated program for molecular replacement. *Journal of Applied Crystallography* 30:1022–1025.
17. Binkowski TA, Joachimiak A (2008) Protein functional surfaces: global shape matching and local spatial alignments of ligand binding sites. *BMC Struct Biol* 8:45–55.
18. Case DA, et al. (2005) The amber biomolecular simulation programs. *J Comput Chem* 26:1668–1688.
19. DeLano WL (2008) PyMOL (Schrödinger, LLC. New York, NY).

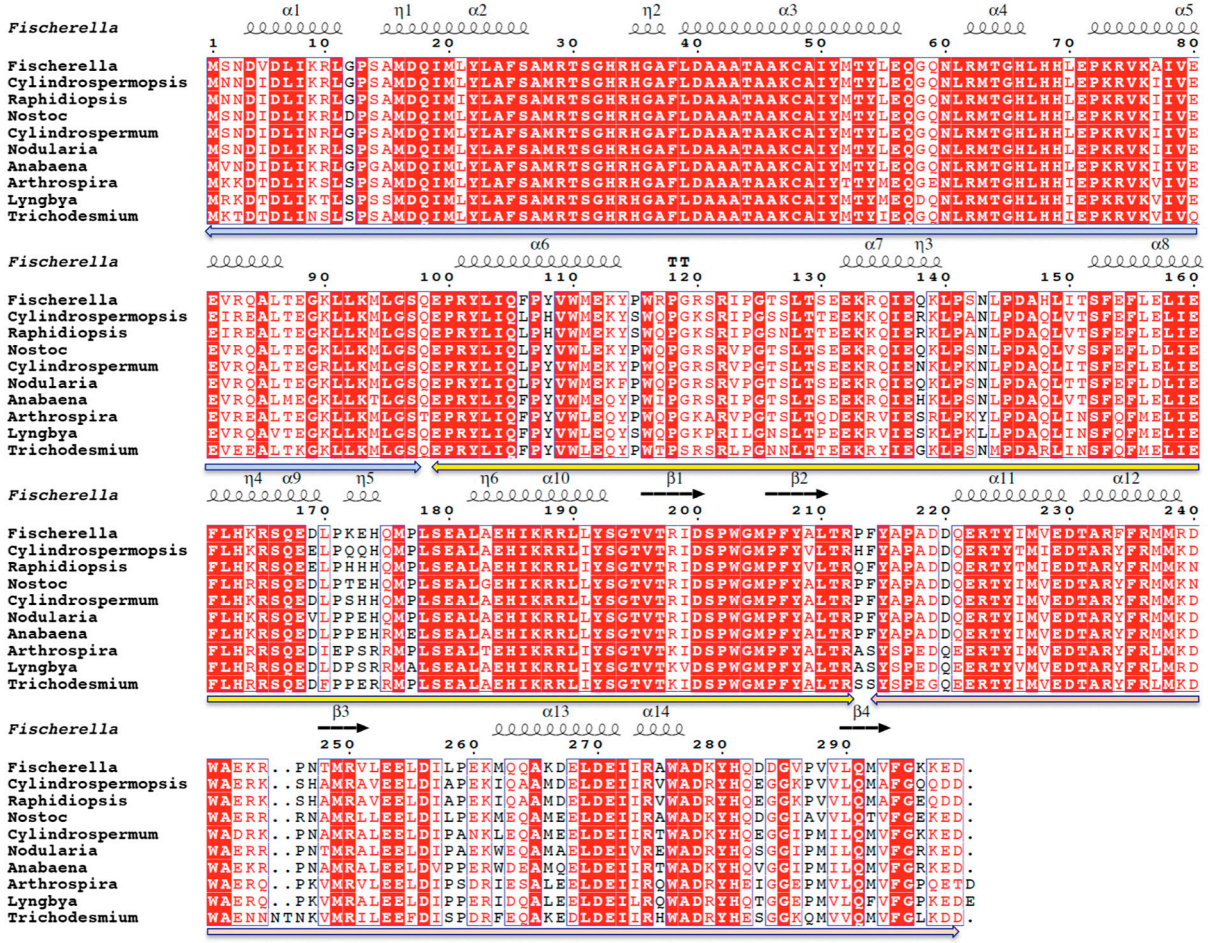


Fig. S1. Multiple sequence alignment of HetR homologs. Sequence identities are highlighted in red and similarities are shown as red letters. The corresponding secondary structure of HetR from *Fischerella* sp. MV11 is shown at the top (black). Helices (α, α-helix; η, 3_{10} helix) appear as small squiggles, beta strands (β, β-strand) as arrows, and a beta turn is labeled TT. The sequences of HetR from the following species were used for alignment (*Fischerella* sp. MV11, *Cylindrospermopsis raciborskii* CS-505, *Raphidiopsis brookii* D9, *Nostoc punctiforme* PCC 73102, *Cylindrospermum* sp. A1345, *Nodularia spumigena* CCY9414, *Anabaena* sp. strain PCC 7120, *Arthrospira maxima* CS-328, *Lyngbya*, sp. PCC 8106, *Trichodesmium erythraeum* IMS101). The figure was prepared with ESPript/ENDscript webserver (1). The highly conserved sequence motifs (listed from N to C terminus) include: residues 16–51 [MDQIMLYLAFSAMRTSGHRHGAFLDAAATAAKCAIY], residues 60–79 [NLRMTGHLHHLEPKRVKAIV], residues 90–105 [KLLKMLGSQEPRYLIQ], residues 158–168 [LIEFLHKRSQE], residues 178–212 [LSEALAEHIKRRLLYSGTVTRIDSPWGMPFYALTR], residues 222–236 [ERTYIMVEDTARFFR], and residues 269–281 [LDEIIRAWADKYH]). The amino acid sequence ranges for the N-terminal domain, the flap domain, and the C-terminal hood domain are shown with blue, yellow, and pink horizontal arrows, respectively.

1 Gouet P, Courcelle E, Stuart DI, Metoz F (1999) ESPript: multiple sequence alignments in PostScript. *Bioinformatics* 15:305–308.

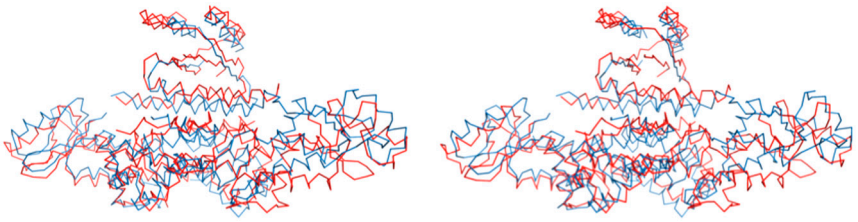


Fig. S2. Stereo view of structural alignment of HetR determined under high ionic strength at 3.38 Å (red) and low ionic strength at 3.0 Å (blue).

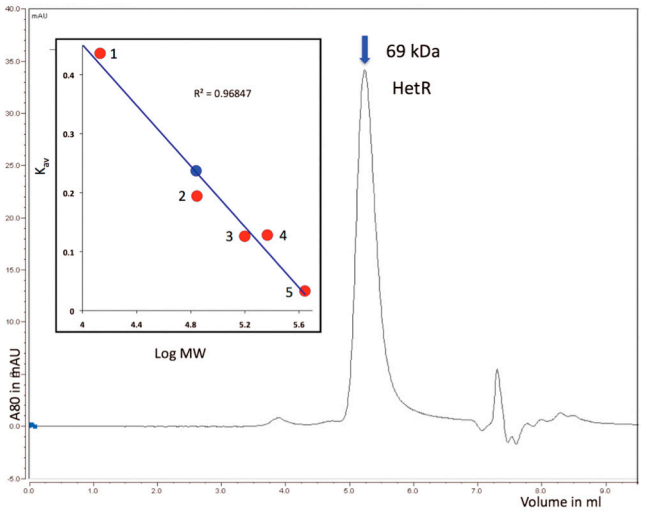


Fig. S3. Size-exclusion chromatography. The absorbance at 280 nm is plotted in absorbance units (AU) vs. retention volume in milliliters. Insert is a plot of Kav coefficient vs. logarithm of molecular weight red circles correspond to standard proteins: (1) ribonuclease (13.5 kDa), (2) conalbumin (75 kDa), (3) aldolase (158 kDa), (4) catalase (232 kDa), and (5) ferritin (440 kDa); blue circle corresponds to HetR (69 kDa). A clear peak is observed for the HetR dimer; no peaks are observed for the monomer or tetramer.

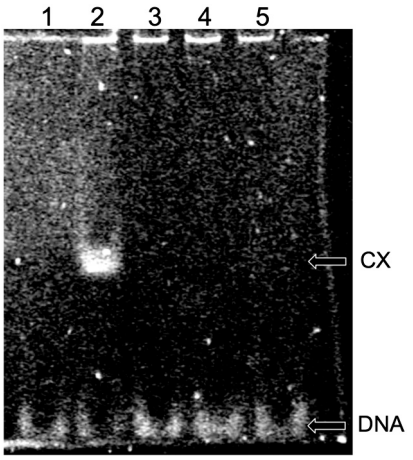


Fig. S4. Lack of DNA binding by mutants of HetR. Samples were separated on agarose gel and stained with SYBR Green. Lanes contain: 1- negative control (H_2O + DNA), 2. HetR (WT) + DNA, 3—HetR (R62E) + DNA, 4—HetR (K73E) + DNA, 5—HetR (R74E) + DNA. Arrows point to position of HetR/DNA complex (CX) and free DNA (DNA). The DNA was the 29-bp palindromic oligonucleotide described by Higa and Callahan (1).

1 Higa KC, Callahan SM (2010) Ectopic expression of HetP can partially bypass the need for HetR in heterocyst differentiation by *Anabaena* sp. strain PCC 7120. *Mol Microbiol* 77:562–574.

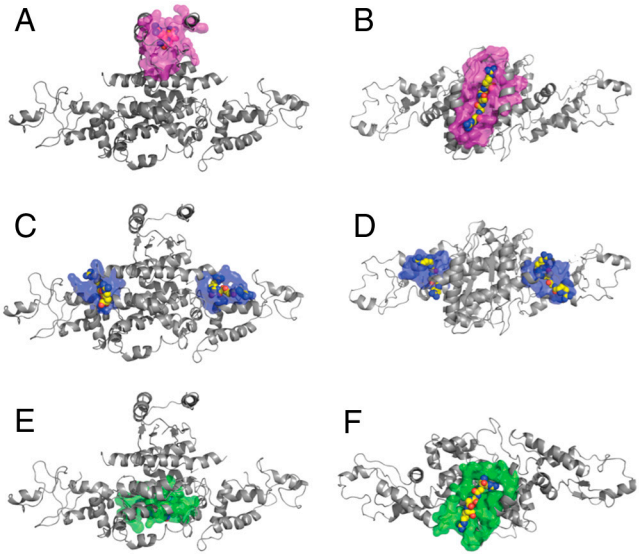
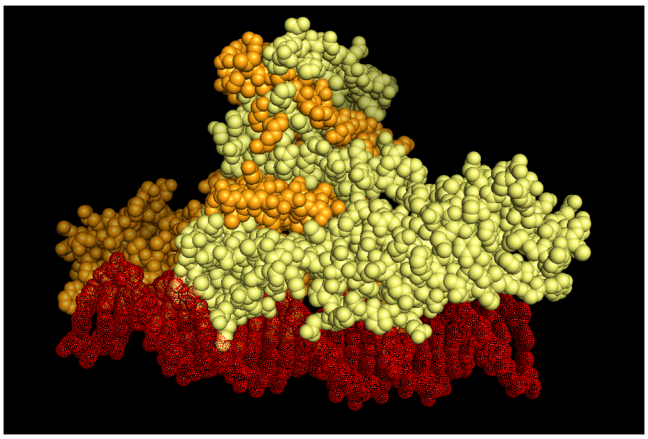


Fig. S5. Potential RGSGR pentapeptide binding sites. HetR surface analysis was performed using GPSS (1).

1 Binkowski TA, Joachimiak A (2008) Protein functional surfaces: global shape matching and local spatial alignments of ligand binding sites. *BMC Struct Biol.* 8:45–55.



Movie S1. PyMol (19) model of the HetR dimer with the 29-bp palindrome bound. The two subunits of HetR are in yellow and tan, respectively; the DNA is red. The three residues shown by mutation to be essential for DNA binding are in green

Movie S1 (MOV)