The *Calyptogena magnifica* Chemoautotrophic Symbiont Genome

I. L. G. Newton,[1] T. Woyke,[2] T. A. Auchtung,[1] G. F. Dilly,[1] R. J. Dutton,[3] M. C. Fisher,[1] K. M. Fontanez,[1] E. Lau,[1] F. J. Stewart,[1] P. M. Richardson,[2] K. W. Barry,[2] E. Saunders,[2] J. C. Detter,[2] D. Wu,[4] J. A. Eisen,[5] C. M. Cavanaugh[1]*

Chemoautotrophic endosymbionts are the metabolic cornerstone of hydrothermal vent communities, providing invertebrate hosts with nearly all of their nutrition. The *Calyptogena magnifica* (Bivalvia: Vesicomyidae) symbiont, *Candidatus Ruthia magnifica*, is the first intracellular sulfur-oxidizing endosymbiont to have its genome sequenced, revealing a suite of metabolic capabilities. The genome encodes major chemoautotrophic pathways as well as pathways for biosynthesis of vitamins, cofactors, and all 20 amino acids required by the clam.

Metazoans at deep-sea hydrothermal vents flourish with the support of symbiotic chemoautotrophic bacteria (*1*). Analogous to photosynthetic chloroplasts, which evolved from cyanobacterial ancestors and use light energy to fix carbon for their plant and algal hosts, chemoautotrophic endosymbionts use the chemical energy of reduced sulfur emanating from vents to provide their hosts with carbon and a large array of additional necessary nutrients. In return, host invertebrates bridge the oxic-anoxic interface to provide symbiotic bacteria with the inorganic substrates necessary for chemoautotrophy. The giant clam, *Calyptogena magnifica* Boss and Turner (Bivalvia: Vesicomyidae), was one of the first organisms described after the discovery of hydrothermal vents (*2*). It has a reduced gut and ciliary food groove and is nutritionally dependent on its γ-proteobacterial symbionts (here named *Candidatus Ruthia magnifica*) (*3*, *4*). We present the complete genome of *Ruthia magnifica* (Fig. 1). Despite having a relatively small genome (1.2 Mb), *R. magnifica* is predicted to encode all the metabolic pathways typical of free-living chemoautotrophs, including carbon fixation, sulfur oxidation, nitrogen assimilation, and amino acid and cofactor/vitamin biosynthesis (fig. S1 and table S1).

The following sections outline the reconstruction of *R. magnifica*'s chemoautotrophic metabolism and what this might mean for the biology of its host. Our analysis provides direct evidence that this symbiont fixes carbon via the Calvin cycle, the dominant form of carbon fixation in vent symbioses (*5*), by using energy derived from sulfur oxidation. The *R. magnifica* genome encodes enzymes specific to this cycle, including a form II ribulose 1,5-bisphosphate carboxylase-oxygenase (RuBisCO) and phosphoribulokinase (*6*). Interestingly, however, it appears that *R. magnifica* lacks homologs of sedoheptulose 1,7-bis phosphatase and fructose 1,6-bis-phosphatase and may regenerate ribulose 1,5-bisphosphate via an unconventional route, one that was a reversible pyrophosphate-dependent phosphofructokinase [supporting online material (SOM) text].

Energy for carbon fixation in *R. magnifica* appears to result from sulfur oxidation via the *sox* (sulfur oxidation) and *dsr* (dissimilatory sulfite reductase) genes (fig. S1). The symbiont may oxidize its sulfur granules via *dsr* homologs when external sulfide is lacking, as occurs in both *Allochromatium vinosum* and *Chlorobium limicola* (*7*, *8*). Homologs encoding both a sulfide:quinone oxidoreductase and rhodanese are also present, and, with the *dsr* and *sox* proteins, these enzymes can oxidize sulfide or thiosulfate to sulfite (fig. S1). Sulfite can then be oxidized to sulfate by adenosine 5′-phosphosulfate (APS) reductase and adenosine triphosphate (ATP) sulfurylase before being exported from the cell via a sulfate transporter. Genomic evidence of the Calvin cycle and the sulfur oxidation pathways confirms the chemoautoautotrophic nature of this symbiosis. These data support prior reports showing RuBisCO and ATP sulfurylase activity in *C. magnifica* gill tissue (*4*, *9*), carbon dioxide uptake by the clam in response to sulfide or thiosulfate (*10*), and sulfide-binding, zinc-containing lipoprotein in the host blood stream (*11*).

Energy conservation via the creation of a charge across a membrane proceeds in *R. magnifica* through a nicotinamide adenine dinucleotide (NADH) dehydrogenase, a sulfide:quinone oxidoreductase, and an *rnf* complex (*12*). The genome encodes a straightforward electron transport chain; thus, the reduced quinone in the symbiont membrane may transfer electrons to cytochrome c via a bc_1 complex, and a terminal cytochrome c oxidase could then transfer these electrons to oxygen (fig. S1).

Our analysis shows that *R. magnifica* has the potential to produce 20 amino acids, 10 vitamins or cofactors, and all necessary biosynthetic intermediates, supporting an essential role of symbiont metabolism in the nutrition of this symbiosis. Two nitrogen assimilation pathways, essential to provisioning of amino acids in the symbiosis, occur in *R. magnifica*. Nitrate and ammonia, which enter the cell via a nitrate or nitrite transporter and two ammonium permeases, are reduced via nitrate and nitrite reductase and assimilated via glutamine synthetase and glutamate synthase, respectively (fig. S1). Although nitrate is the dominant nitrogen form at vents (*13*), the symbiont may also assimilate ammonia via recycling of the host's amino acid waste. Unlike any other sequenced endosymbiont genome, *R. magnifica* encodes complete pathways for the biosynthesis of 20 amino acids, including 9 essential amino acids or their precursors (fig. S3). *R. magnifica* also has complete biosynthetic pathways for the majority of vitamins and cofactors (SOM text). The genome encodes a complete glycolytic pathway and the nonoxidative branch of the pentose phosphate pathway and, interestingly, also encodes a tricarboxylic acid (TCA) cycle lacking α-ketoglutarate dehydrogenase. The lack of this enzyme has been suggested to indicate obligate autotrophy in other bacteria (*14*). Carbon fixed via the Calvin cycle can enter the TCA cycle through phosphoenolpyruvate and here could follow biosynthetic routes either to fumarate or to α-ketoglutarate.

As with other intracellular species, genes not found in the *R. magnifica* genome reveal important details about the interaction between host and symbiont. Few substrate-specific transporters were found, suggesting that the symbionts are leaky or that the host actively digests symbiont cells. Indeed, *Ruthia*'s closest known relatives, the bathymodiolid mussel symbionts, are digested intracellularly by their host (*15*). Although the vesicomyids and the bathymodiolids are distantly related, putative degradative stages of symbionts also are observed within *C. magnifica* bacteriocytes (fig. S2b). The symbiont is also lacking the key cell division protein *ftsZ* suggesting that, like *Chlamydia* spp., intracellular division may not proceed through usual pathways (SOM text).

Intracellular endosymbionts often follow distinctive evolutionary routes, including genome reduction, skewed base compositions, and elevated mutation rates (*16*). Given the apparent defects in DNA repair in *R. magnifica* (SOM text) and the likely evolutionary forces pushing this genome toward degradation, it is particularly informative that *Ruthia* retains genes encoding a full suite of chemoautotrophic processes. Indeed, *R. magnifica* has the largest genome of any intracellular symbiont sequenced to date and may represent an early intermediate in the evolution toward a plastid-like chemoautotrophic organelle. The broad array of metabolic pathways revealed through sequencing of the *R. magnifica* genome confirms and extends prior knowledge of host nutritional dependency.

[1]Harvard University, 16 Divinity Avenue, Biolabs 4080, Cambridge, MA 02138, USA. [2]Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA. [3]Harvard Medical School, Department of Microbiology and Molecular Genetics, 200 Longwood Avenue, Boston, MA 02115, USA. [4]Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA. [5]University of California, Davis Genome Center, Genome and Biomedical Sciences Facility, Room 5311, 451 East Health Sciences Drive, Davis, CA 95616–8816, USA

*To whom correspondence should be addressed. E-mail: cavanaug@fas.harvard.edu
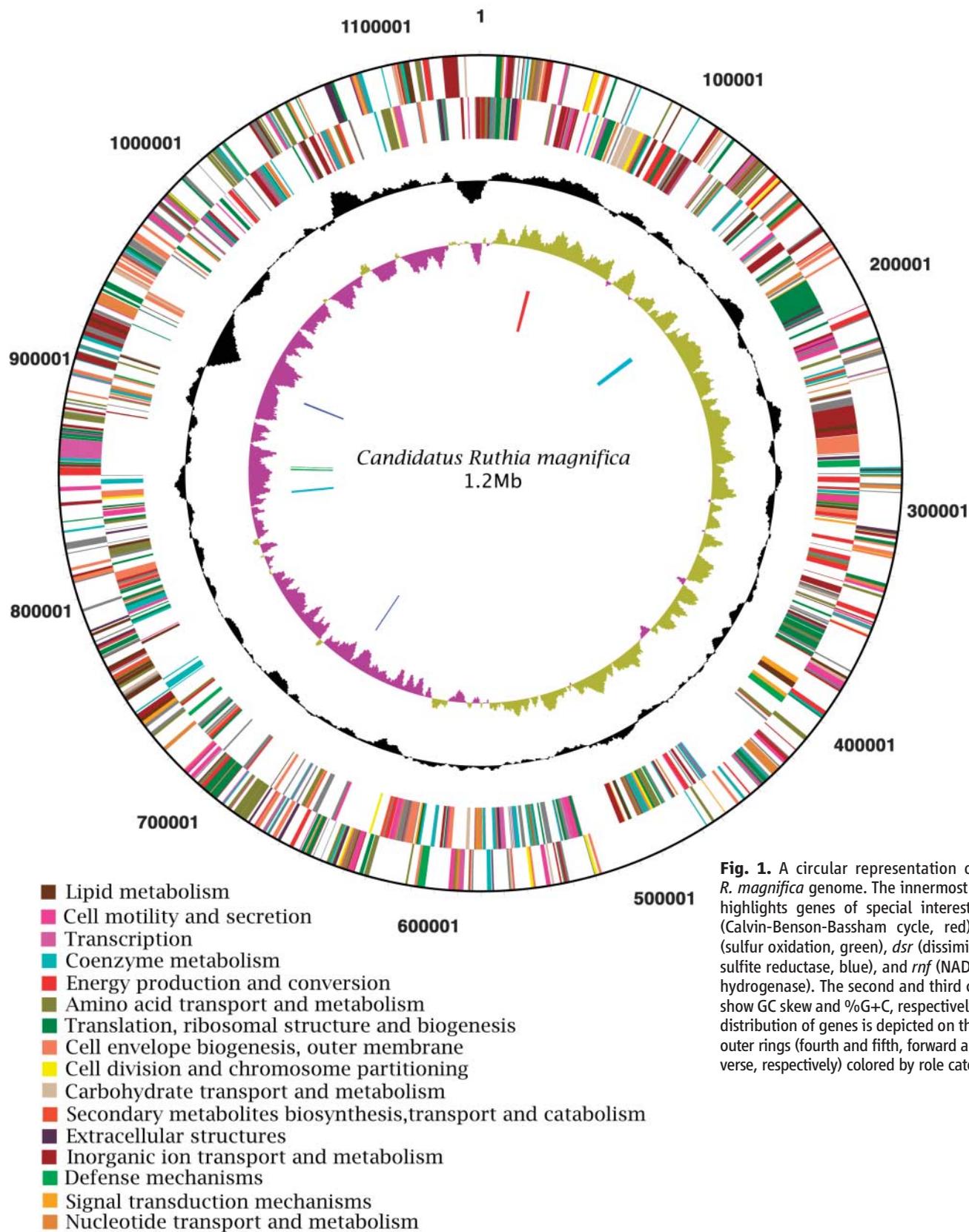
Fig. 1. A circular representation of the *R. magnifica* genome. The innermost circle highlights genes of special interest: *cbb* (Calvin-Benson-Bassham cycle, red), *sox* (sulfur oxidation, green), *dsr* (dissimilatory sulfite reductase, blue), and *rnf* (NADH dehydrogenase). The second and third circles show GC skew and %G+C, respectively. The distribution of genes is depicted on the two outer rings (fourth and fifth, forward and reverse, respectively) colored by role category.

Legend (by color):
- Lipid metabolism
- Cell motility and secretion
- Transcription
- Coenzyme metabolism
- Energy production and conversion
- Amino acid transport and metabolism
- Translation, ribosomal structure and biogenesis
- Cell envelope biogenesis, outer membrane
- Cell division and chromosome partitioning
- Carbohydrate transport and metabolism
- Secondary metabolites biosynthesis, transport and catabolism
- Extracellular structures
- Inorganic ion transport and metabolism
- Defense mechanisms
- Signal transduction mechanisms
- Nucleotide transport and metabolism

References and Notes

1. C. M. Cavanaugh, Z. P. McKiness, I. L. G. Newton, F. J. Stewart, in *The Prokaryotes* (Springer-Verlag, Berlin, 2004).
2. K. J. Boss, R. D. Turner, *Malacologia* **20**, 161 (1980).
3. C. M. Cavanaugh, *Nature* **302**, 58 (1983).
4. J. J. Robinson, thesis, Harvard University, Cambridge, MA (1997).
5. C. M. Cavanaugh, J. J. Robinson, in *Microbial Growth on C₁ Compounds*, M. E. Lidstrom, F. R. Tabita, Eds. (Kluwer Academic, Dordrecht, Netherlands, 1996), pp. 285–292.
6. J. M. Shively, G. Van Keulen, W. G. Meijer, *Annu. Rev. Microbiol.* **52**, 191 (1998).
7. C. Dahl *et al.*, *J. Bacteriol.* **187**, 1392 (2005).
8. F. Verte *et al.*, *Biochemistry* **41**, 2932 (2002).

9. C. R. Fisher *et al.*, *Deep-Sea Res.* **35**, 1811 (1988).
10. J. J. Childress, C. R. Fisher, J. A. Favuzzi, N. K. Sanders, *Physiol. Zool.* **64**, 1444 (1991).
11. F. Zal *et al.*, *Cah. Biol. Mar.* **41**, 413 (2000).
12. H. Kumagai, T. Fujiwara, H. Matsubara, K. Saeki, *Biochemistry* **36**, 5509 (1997).
13. K. S. Johnson, J. J. Childress, R. R. Hessler, C. M. Sakamoto-Arnold, C. L. Beehler, *Deep-Sea Res.* **35**, 1723 (1988).
14. A. P. Wood, J. P. Aurikko, D. P. Kelly, *FEMS Microbiol. Rev.* **28**, 335 (2004).
15. A. Fiala-Médioni, C. Métivier, A. Herry, M. Le Pennec, *Mar. Biol.* **92**, 65 (1986).
16. J. J. Wernegreen, *Nat. Rev. Genet.* **3**, 850 (2002).
17. This research was funded by grants from the U.S. Department of Energy/Joint Genome Institute (JGI) Microbial Genomes Program (to C.M.C. and J.A.E.), a Merck Genome-Related Research Award (to C.M.C.), and a Howard Hughes Medical Institute Predoctoral Fellowship (to I.L.G.N.). Sequencing was carried out at the JGI, and we thank D. Bruce and E. Rubin for project management, D. Fraenkel for review of metabolic pathways in *Ruthia*, and P. Girguis and K. Scott for their helpful comments on the manuscript. The *Candidatus R. magnifica* genome can be found in GenBank (CP000488).

The Phosphothreonine Lyase Activity of a Bacterial Type III Effector Family

Hongtao Li,* Hao Xu,* Yan Zhou,* Jie Zhang, Chengzu Long, Shuqin Li, She Chen, Jian-Min Zhou, Feng Shao†

Pathogenic bacteria use the type III secretion system to deliver effector proteins into host cells to modulate the host signaling pathways. In this study, the *Shigella* type III effector OspF was shown to inactivate mitogen-activated protein kinases (MAPKs) [extracellular signal–regulated kinases 1 and 2 (Erk1/2), c-Jun N-terminal kinase, and p38]. OspF irreversibly removed phosphate groups from the phosphothreonine but not from the phosphotyrosine residue in the activation loop of MAPKs. Mass spectrometry revealed a mass loss of 98 daltons in p-Erk2, due to the abstraction of the α proton concomitant with cleavage of the C-OP bond in the phosphothreonine residue. This unexpected enzymatic activity, termed phosphothreonine lyase, appeared specific for MAPKs and was shared by other OspF family members.

Gram-negative bacterial pathogens often use the type III secretion system (TTSS) to inject into host cells effector proteins (*1*) that interfere with host signal transduction pathways to promote pathogen infection. MAPK signaling plays an important role in activating host innate immune responses in both plants and animals and is a frequent target of pathogenic effectors (*2–4*). Identification of host targets of TTSS effectors and elucidation of their biochemical functions are critical in understanding the mechanism and the evolution of bacterial pathogenesis (*3, 5, 6*).

The TTSS effector OspF is present in all the four known pathogenic species of *Shigella* (*7–9*), a causative agent of bacillary dysentery (*10*). OspF, together with SpvC and HopAI1, represents a family of TTSS effectors conserved in both plant and animal bacterial pathogens (*11*) (fig. S1). None of the OspF family effectors has an established biological or biochemical function. To define the biochemical function of OspF and the OspF family of effectors as well as to identify their potential in vivo targets, we examined effects of OspF on the host immune signaling pathways, including those of nuclear factor κB (NFκB) and MAPKs.

Transient expression of OspF in human 293T cells efficiently blocked extracellular signal–regulated kinases 1 and 2 (Erk1/2) and c-Jun N-terminal kinase (JNK) signaling, whereas it did not alter NFκB activation (Fig. 1A). OspF abrogated phosphorylation of endogenous Erk1/2, JNK, and p38 kinases (Fig. 1, B and C). These results suggest that OspF harbors a specific activity of down-regulating multiple MAPKs but not NFκB signaling when overexpressed in mammalian cells. To define the specific step(s) in the MAPK pathway blocked by OspF, we turned to the canonical Erk pathway that is activated by a phosphorylation cascade of Raf, MAPK kinase (MEK), and Erk kinases (*12*). OspF blocked Erk activation in 293T cells transfected with either RasV12, v-Raf (constitutive active Raf), or MEK1-ED (constitutive active MEK1) (Fig. 1D). Meanwhile, RasV12-induced MEK1 phosphorylation remained intact despite the disappearance of Erk1/2 phosphorylation in cells expressing OspF (Fig. 1E). Thus, inhibition of Erk phosphorylation by OspF is downstream of MEK1 along the classical Ras-Raf-MEK-Erk cascade.

Erk1/2 phosphorylation can be reconstituted in the cell-free extracts system by adding upstream signaling molecules (*3*). Addition of bacterially expressed and highly purified OspF (fig. S2A) in 293T cell extracts abolished both RasV12 and active B-Raf–induced Erk1/2 phosphorylation, whereas MEK1 phosphorylation was not affected (Fig. 2A). This excludes the transcriptional regulation of MAPK pathway by OspF and also provides a system that recapitulates the inhibition of MAPK signaling by OspF in cells. To further test whether OspF directly targets the Raf-MEK1-Erk1/2 cascade biochemically, we reconstituted the phosphorylation cascade of B-Raf–MEK1–Erk2 by using purified kinases. Addition of OspF to this reaction largely abolished Erk2 phosphorylation (Fig. 2B). Furthermore, recombinant OspF could inhibit phosphorylation of glutathione *S*-transferase (GST)–Erk2 by MEK1-ED (Fig. 2C), suggesting that OspF directly targets Erks or MEKs. Recombinant Erk2 fused to maltose binding protein (MBP-Erk2), but not MBP alone, effectively precipitated OspF (fig. S2A). Flag-OspF also coprecipitated with endogenous Erk1/2 but not MEK1 from 293T cells (fig. S2B). Thus, OspF targets Erk and attenuates its phosphorylation status.

Given the direct association between OspF and Erk2 and that the basal amount of phospho-Erk1/2 (p-Erk1/2) was diminished by OspF addition in the cell-free reconstitution assay (Fig. 2A), we reasoned that OspF could harbor a phosphatase activity to reverse Erk phosphorylation. When in vitro phosphorylated GST-Erk2, JNK, and p38 were incubated with recombinant OspF, dephosphorylation indeed occurred (Fig. 2, D and E). Furthermore, OspF selectively removed phosphate groups from the threonine but not from the tyrosine residue in the pT-X-pY (*13*) motif in p-Erk2 (Fig. 2, D and F). Thus, OspF harbors an in vitro enzymatic activity of specifically removing phosphate groups from the phosphothreonine in the pT-X-pY motif in MAPKs, including Erk1/2, JNK, and p38.

OspF appears to function as a threonine-specific MAPK phosphatase. However, our analysis rules out the possibility of a classical protein phosphatase of OspF in nature [Supporting Online Material (SOM) text]. To reveal the chemical nature of OspF activity, we subjected MEK1-phosphorylated MBP-Erk2 (p-Erk2–control), as well as MBP–p-Erk2 further treated with OspF (p-Erk2–OspF), to tandem mass spectrometry analysis. More than 80% of the MBP-Erk2 sequence was covered by observed tandem peptides. All of the peptides from p-Erk2–control and p-Erk2–OspF were indistinguishable except for the peptide containing the TXY phosphomotif [VADPDHDHTGFL-pT-E-pY-VATR (*13*)]. The mass of the peptide from p-Erk2–OspF was 2204.9 dalton (Fig. 3A), 98 dalton less than that of the corresponding peptide from p-Erk2–control. The 98-dalton mass decrease occurred on the phosphothreonine residue in the pT-X-pY

National Institute of Biological Sciences, Beijing, 102206, China.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: shaofeng@nibs.ac.cn