

## Supplemental Material

### Structural and biochemical analysis reveals a distinct catalytic site of salicylate 5-monooxygenase NagGH from Rieske dioxygenases

Yan-Jie Hou<sup>1,3#</sup>, Yuan Guo<sup>2#</sup>, De-Fen Li<sup>1,3\*</sup>, Ning-Yi Zhou<sup>2\*</sup>

<sup>1</sup>State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

<sup>2</sup>State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic and Developmental Sciences, and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200240, China.

<sup>3</sup>National Laboratory of Biomacromolecules, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

Corresponding authors: De-Feng Li (Email: [lidefeng@im.ac.cn](mailto:lidefeng@im.ac.cn)) and Ning-Yi Zhou (Email: [ningyi.zhou@sjtu.edu.cn](mailto:ningyi.zhou@sjtu.edu.cn))

# These authors contributed equally to this work

Running title: Crystal structure of salicylate 5-monooxygenase NagGH

**Abbreviations:** Rieske non-heme iron oxygenase (RO), biphenyl dioxygenase (BPDO), naphthalene dioxygenase (NDO), nitrobenzene dioxygenase (NBDO), polycyclic aromatic hydrocarbons (PAH).

**Keywords:** polycyclic aromatic hydrocarbons degradation, salicylate 5-monooxygenase, crystal structure, product regioselectivity, Rieske non-heme iron oxygenase

**TABLE S1 Crystallographic data collection and model refinement statistics**

NagGH	
<b>Data collection</b>	
Space group	P6 <sub>3</sub>
Cell dimensions	
<i>a, b, c</i> (Å)	188.16, 188.16, 147.59
$\alpha, \beta, \gamma$ (°)	90, 90, 120
Resolution (Å)	47.28-2.60 (2.64-2.60)*
No. reflections	91061 (4515)
<i>R</i> <sub>pim</sub>	0.067 (0.404)
<i>I</i> / $\sigma I$	12.3 (2.2)
Completeness (%)	100.0 (100.0)
Redundancy	6.4 (6.4)
CC(1/2)	0.994 (0.421)
<b>Refinement</b>	
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.191 / 0.235
No. of Non-H atoms	
Total	18150
Protein	17804
Ligand/ion	20
Water	326
<i>B</i> -factors	
Total	51.05
Protein	51.18
Ligand/ion	50.39
Water	43.73
R.m.s. deviations	
Bond lengths (Å)	0.005
Bond angles (°)	1.160
Ramachandran favored (%)	95.48
Ramachandran allowed (%)	4.34
Ramachandran outliers (%)	0.19

\*Values in parentheses are for highest-resolution shell.

```

1      10      20      30      40      50      60      70      80
NagG   MSEPQRLKPVFPQDPKWP...G.....EGSSRVFFWYATREDLYKRLELRFYANHWCYVGLAEINPGRFVRRVIGERSVIMVRDPE
NDO9816-4 .....MMYNN...KILVSSGLSQKHLIHGDEELFQHELKITIFARNWFLTHDSLIPAPGDYVVTAKMGIDEVIVSRQND
BPDOLB400 MS..SAIKEVQGAPVVKWVTNWTPEAIRGLVDOEKGLLDPRLYADQSLYELERLRFYANHWCYVGLAEINPGRFVRRVIGERSVIMVRDPE

90      100      110      120      130      140      150      160      170
NagG   GGIVNVEVCAHRCMRFCRERHGNAKDFPCYHWNYSGLKGDIQGVFFRRCVYKQDQKVNNGMPKDFKLEEHGLTKLVVAARGAVFASDHDVE
NDO9816-4 GSTRAFLNVCRRGKTLVSVVEAGNAGGFCVSYHGWGFGSNGBLIQSVFFEKDLYGKESLNKCC...LGL...KEVARVTSFHFGLIYGCDEQEAPE
BPDOLB400 KSIKVFLLNQCRRHRCMRICRSDAAGNAGGFCVSYHGWGFGSNGBLIQSVFFEKDLYGKESLNKCC...LGL...KEVARVTSFHFGLIYGCDEQEAPE

180      190      200      210      220      230      240      250      260
NagG   PFEEFLCPTILHYFDRVFN...GRKLLTLCYRQRITPNNKLMQENIKDPPYHPCLLHTWFSTFGLWRADNKSELKMDAKFRHAMITSTRCQGG
NDO9816-4 PLMDYILCDAAWYLEMFKHS.GGLELVGFPQKVVITKANWKPAPENFVGDYHVTWTHASS.LR...SG...ESIFSSLAGNAALPPECAGL
BPDOLB400 DLETYILCDAR.PYMDVMLDRTPAGTVAIGCMQKVVITPCNWKFAADQFCSDMYHACITTHLSGIL...AG...IPPEDM.LSQAQITPKG...

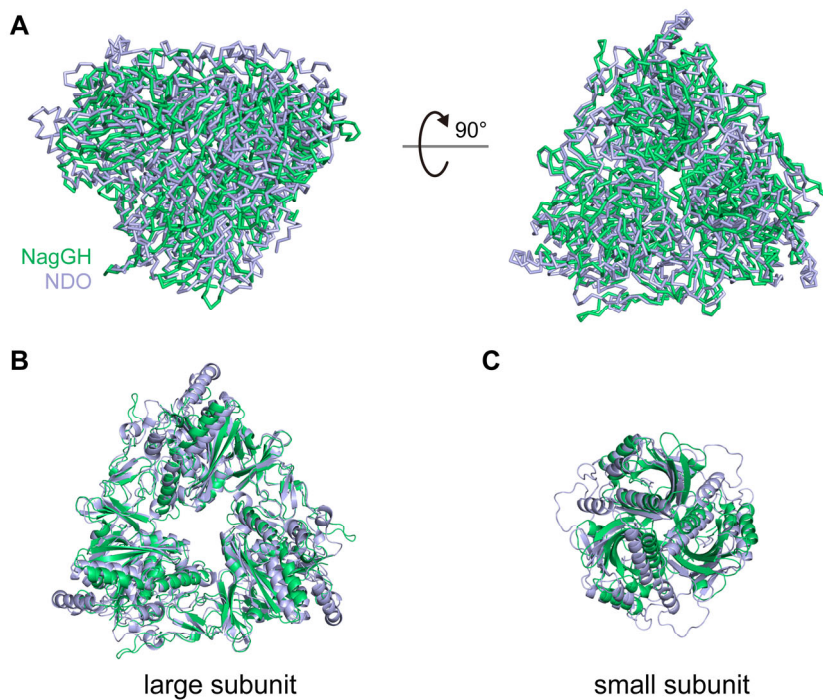
270      280      290      300      310      320
NagG   KNEE.....VVSQVDSFKEQMKVNDPRLLDIVPEPWWGGPTA.....VMTTIFPSVVIQQQVNSVSTRHIO
NDO9816-4 QMTSKYSGSMGLWDYSGVHSAADLVPELMAFGAK.....QERLNKEIGDVRARIY.RSHLNCVTFPNNNSMLTCSGVFKV..WN
BPDOLB400 .....NQFRAAWGGHSGWYVDEPGLLAVMGPKV...TQYWTGEPAAELAEQRLGHTGMPVRRMVGQHMVTFPNCVSLPTFNNIRI..WH

330      340      350      360      370      380      390      400
NagG   PNGTHSDFVWTHFGREDNEEWQRRLL.IQANLFGPAQFVSAADDGNEVIEWSQEGFEQKPTHR...TVTEMCGHEICDID.....HMTB
NDO9816-4 PIDAN.TTEVWYAIVEKDMPEDLKRRLLADSVQRTFGPAQFVSAADDGNEVIEWSQEGFEQKPTHR...TVTEMCGHEICDID.....HMTB
BPDOLB400 PRGPN.EIEVWAFITLVADAPAEIKEEYRRRNIRNFSAGQVFEQDDGNEVIEWSQEGFEQKPTHR...TVTEMCGHEICDID.....HMTB

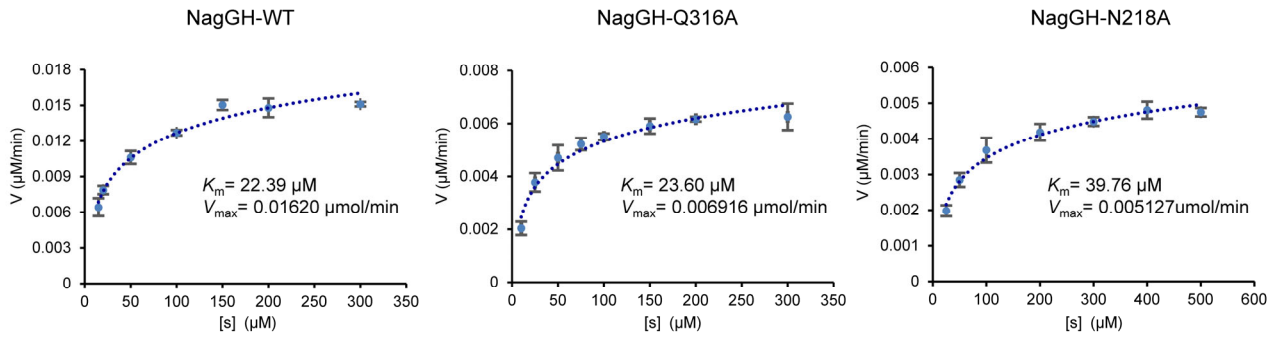
410      420
NagG   TLIKRMVDYWRKVMGE.....
NDO9816-4 TSYRGRFYRAYQAHVSSSNWAEFEHASSTWHTLTKTTDR
BPDOLB400 EAAKRMVHHWRRMMSSEPSWATLKP.....

```

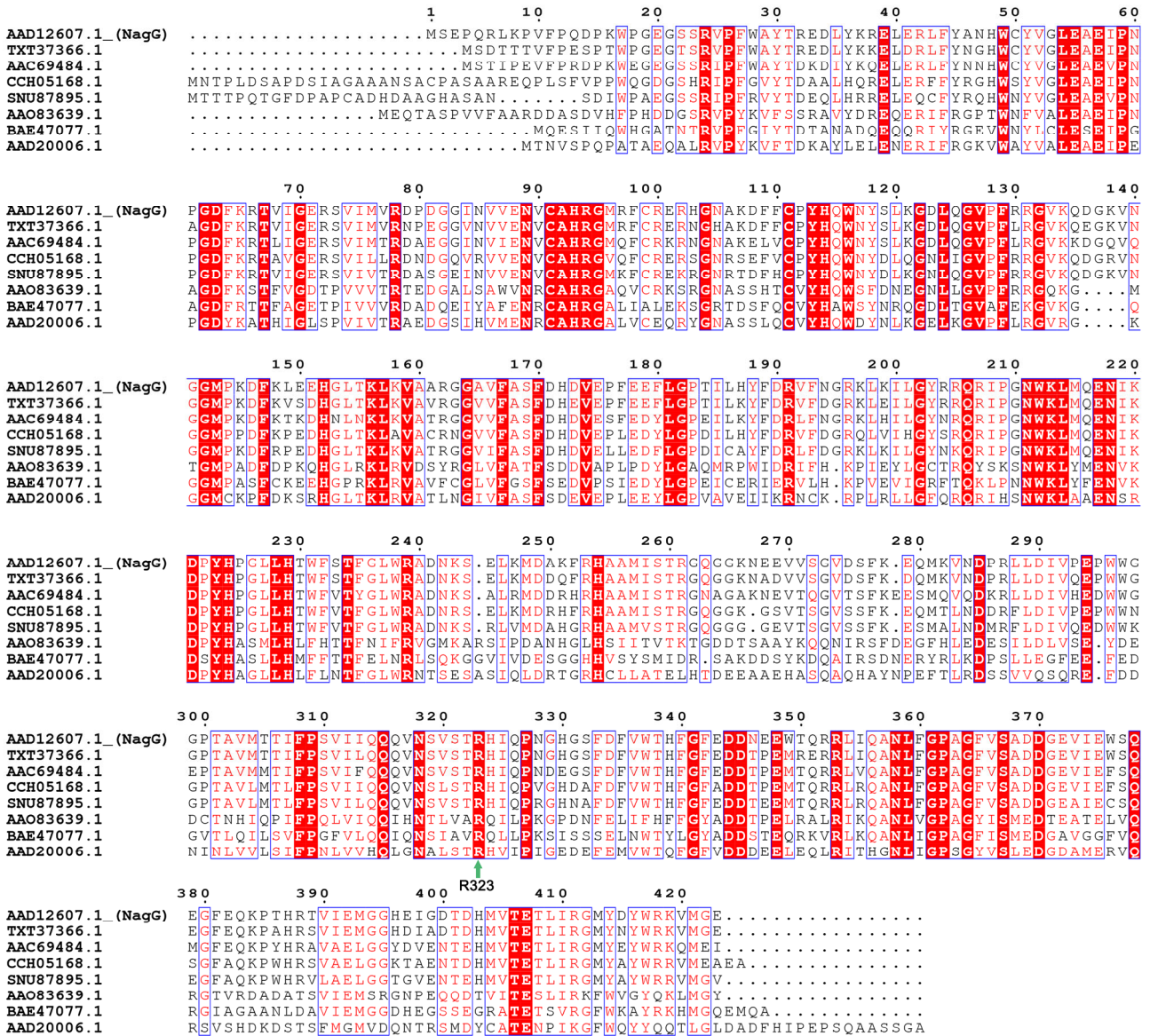
**FIG S1** Sequence alignment of NagG, the large subunits of NDO<sub>9816-4</sub> and BPDO<sub>LB400</sub>. Sequence number of NagG is labelled on the top of protein sequences.



**FIG S2** The superposition of NagGH and NDO. The overall superposition of NagGH (green) and NDO<sub>9816-4</sub> (grey) hexamers are shown in (A) whereas the respective large (NagG) and small (NagH) subunit superposition are in (B) and (C).



**Fig S3** Enzymatic assays of NagGH. Enzymatic assays of NagGH and its mutants. The kinetic curves of NagGH-WT, NagGH-Q316A and NagGH-N218A are shown here. The x-coordinate shows the substrate concentration, and the y-coordinate shows the rate of reaction.



**FIG S4** Sequence alignment of NagG and aromatic acid dioxygenases. AAD12607.1 (NagG): this study. TXT37366.1: cinnamic acid dioxygenase alpha subunit from *Comamonadaceae* bacterium; AAC69484.1: 2-hydroxybenzoate 5-hydroxylase alpha subunit from *Pseudomonas aeruginosa*; CCH05168.1: Ortho-halobenzoate 1,2-dioxygenase alpha subunit from *Achromobacter xylosoxidans* NH44784-1996; SNU87895.1: Anthranilate 1,2-dioxygenase large subunit from *Pandoraea sputorum*; AAO83639.1: Anthranilate 1,2-dioxygenase large subunit from *Burkholderia cepacian*; BAE47077.1: Terephthalate 1,2-dioxygenase from *Comamonas* sp.; AAD20006.1: Ortho-halobenzoate 1,2-dioxygenase from *Pseudomonas aeruginosa*. Sequence number of NagG is labelled on the top of protein sequences.