

**Manuscript version: Author's Accepted Manuscript**

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

**Persistent WRAP URL:**

<http://wrap.warwick.ac.uk/148309>

**How to cite:**

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Publisher's statement:**

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk).

# Molecular basis for control of antibiotic production by a bacterial hormone

Shanshan Zhou,<sup>1†</sup> Hussain Bhukya,<sup>4†</sup> Nicolas Malet,<sup>1†</sup> Peter J. Harrison,<sup>1</sup> Dean Rea,<sup>2</sup> Matthew J. Belousoff,<sup>5</sup> Hariprasad Venugopal,<sup>6</sup> Paulina K. Sydor,<sup>1</sup> Kathryn M. Styles,<sup>2</sup> Lijiang Song,<sup>1</sup> Max J. Cryle,<sup>4,7</sup> Lona M. Alkhalaf,<sup>1</sup> Vilmos Fülöp,<sup>2</sup> Gregory L. Challis<sup>1,3,4,7\*</sup> and Christophe Corre<sup>1,2,3\*</sup>

<sup>1</sup>Department of Chemistry, University of Warwick, Coventry CV4 7AL, UK

<sup>2</sup>School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK

<sup>3</sup>Warwick Integrative Synthetic Biology Centre, University of Warwick, Coventry CV4 7AL, UK

<sup>4</sup>Biomedicine Discovery Institute, Department of Biochemistry and Molecular Biology, Monash University, Clayton VIC 3800, Australia.

<sup>5</sup>Infection and Immunity Program, Department of Microbiology, Monash University, Clayton, VIC 3800, Australia.

<sup>6</sup>Ramaciotti Centre for Electron Microscopy, Monash University, Clayton, VIC 3800, Australia.

<sup>7</sup>ARC Centre of Excellence for Innovations in Peptide and Protein Chemistry, Monash University, Clayton VIC 3800, Australia.

† These authors contributed equally

\*Corresponding authors:

Dr C. Corre, School of Life Sciences, University of Warwick, Coventry CV4 7AL, U.K. Tel 44-2476-523557. Email [C.Corre@warwick.ac.uk](mailto:C.Corre@warwick.ac.uk)

Prof G.L. Challis, Department of Chemistry, University of Warwick, Coventry CV4 7AL, U.K. Tel 44-2476-574024. Email [G.L.Challis@warwick.ac.uk](mailto:G.L.Challis@warwick.ac.uk)



## Summary

Actinobacteria produce numerous antibiotics and other specialised metabolites with important applications in medicine and agriculture<sup>1</sup>. Diffusible hormones frequently control the production of such metabolites by binding TetR family transcriptional repressors (TFTRs), but the molecular basis for this remains unclear<sup>2</sup>. The production of methylenomycin antibiotics in *Streptomyces coelicolor* A3(2) is initiated by binding of 2-alkyl-4-hydroxymethylfuran-3-carboxylic acid (AHFCA) hormones to the TFTR MmfR<sup>3</sup>. Here, we report the X-ray crystal structure of an MmfR-AHFCA complex, establishing the structural basis for hormone recognition. We also elucidate the mechanism for DNA release upon hormone binding by single particle cryo-electron microscopy of an MmfR-operator complex. DNA binding and release assays with MmfR mutants and synthetic AHFCA analogues illuminate the role played by individual amino acid residues and hormone functional groups in ligand recognition and DNA release. These findings will facilitate the exploitation of Actinobacterial hormones and their associated TFTRs in synthetic biology and novel antibiotic discovery.

## Introduction

Actinobacteria typically have a complex life cycle that proceeds from spore germination, through branched multi-nucleoid hyphae, to aerial hyphae, which septate into mono-nucleoid compartments that become spores<sup>4</sup>. Specialised metabolite production is coordinated with this cycle, usually commencing at the onset of aerial growth<sup>5</sup>.

Diffusible hormones frequently induce the expression of specialised metabolite biosynthetic gene clusters (BGCs) in Actinobacteria<sup>6</sup>. The archetypal example is A-factor, a  $\gamma$ -butyrolactone (GBL) that triggers aerial mycelium formation and production of the antibiotic streptomycin in *Streptomyces griseus* (Figure 1a)<sup>7</sup>. Binding of A-factor to ArpA, a TetR family transcriptional repressor (TFTR), releases it from the promoter of a transcriptional activator that induces the expression of genes controlling morphogenesis and antibiotic biosynthesis (Figure 1b)<sup>7</sup>.

Antibiotic production (and in some cases morphogenesis) is controlled by analogous mechanisms in several other Actinobacteria. For many years, GBLs were believed to be only hormones involved. However, over the past decade three additional hormone classes have been implicated in the induction of antibiotic biosynthesis via binding to ArpA-like TFTRs (Figure 1a)<sup>3,8,9</sup>. Moreover, such TFTRs regulate the biosynthesis of several commercially important metabolites, but the hormones these respond to are mostly unknown<sup>2</sup>.

Methylenomycin A is an antibiotic produced by *Streptomyces coelicolor* A3(2) (Figure 1c)<sup>10</sup>. We previously reported that a group of five AHFCAs called the methylenomycin furans (MMFs) induce the production of methylenomycin A in *S.*

*coelicolor*<sup>3</sup>. A three-gene operon (*mmfLHP*) at one end of the methylenomycin BGC directs MMF biosynthesis (Figure 1c)<sup>11</sup>. The divergent *mmfR* gene upstream of this operon encodes an ArpA-like TFTR that is hypothesized to bind the *mmfR-mmfl*, and *mmyB-mmyY* intergenic regions<sup>11</sup>. Binding of the MMFs to MmfR is proposed to release it from these regions, allowing *mmyB*, which encodes an activator of the methylenomycin biosynthetic genes, to be expressed<sup>11</sup>.

The mechanisms of ligand recognition and DNA release by TFTRs that regulate antibiotic resistance gene expression in Actinobacteria are well characterised (Extended Data Figure 1a)<sup>12-14</sup>. However, the architecture of ArpA-like TFTRs and their DNA complexes differs significantly from these<sup>15,16</sup>, and the molecular basis for hormone recognition and DNA release are poorly understood. Here we report structures of MmfR bound to an AHFCA and the operator from the *mmfl-mmfR* intergenic region, shedding light on hormone recognition and the mechanism of DNA release. We also report DNA binding and release assays employing wild type and mutant MmfR proteins, as well as a synthetic library of naturally occurring AHFCAs and analogues, which illuminate the role played by key amino acid residues and hormone functional groups in ligand recognition and DNA release.

## Results and Discussion

### MmfR DNA binding and release by MMFs

Electrophoretic Mobility Shift Assays (EMSAs) with purified recombinant MmfR (Extended Data Figure 2a) showed that it binds the *mmfl-mmfR* and *mmyB-mmyY* intergenic regions (Extended Data Figures 3a, 3b). Bioinformatics analyses identified homologous 18 bp pseudo-palindromic operator sequences,

hypothesised to be methylenomycin furan-autoregulator responsive elements (MAREs) in each of these intergenic regions<sup>11</sup>. Fluorescence anisotropy (FA) measurements with DNA duplexes containing MARE1 and MARE2 established that MmfR binds to both sequences (Extended Data Figure 4a). Subtle differences, in particular a terminal 5'-AAA...TTT-3' sequence in MARE2 versus a 5'-ATA...TAT-3' sequence in MARE1, may explain the difference in affinity.

The ability of each MMF produced by *S. coelicolor* to promote MmfR release from the *mmfL-mmfr* and *mmvB-mmvy* intergenic regions was confirmed using EMSAs (Extended Data Figures 3a, 3b and 4b). FA measurements determined the concentration of each hormone required for half maximal release (EC<sub>50</sub>) of MmfR from MARE1 and MARE2 (Extended Data Figure 4b; Supplementary Figures 2 and 3). We also investigated the minimum quantity of each MMF needed to trigger methylenomycin production in *S. coelicolor* (Extended Data Figure 2b and Extended Data Table 1)<sup>3</sup>.

In addition to AHFCAs, *S. coelicolor* produces GBLs that control expression of the coelimycin BGC by binding to the MmfR homologue ScbR<sup>17</sup>. To investigate whether MmfR is specific for AHFCAs or is also able to respond to other classes of hormone, we synthesised SCB1 (Extended Data Figure 5a), an abundant GBL in *S. coelicolor*. EMSAs showed that SCB1 cannot dissociate MmfR from the *mmfL-mmfr* intergenic region (Extended Data Figure 3c), indicating no crosstalk between the AHFCA and GBL-dependent regulation systems in *S. coelicolor*.

### **Crystal structure of MmfR-MMF2 complex**

The X-ray crystal structures of MmfR and an MmfR-MMF2 complex were solved at 1.5 Å resolution (Figure 2a; Extended Data Figure 6a). MmfR has a similar overall

fold to the putative GBL-dependent TFTR CprB (Extended Data Figure 6b, c). The first three  $\alpha$ -helices form the DNA-binding domain (DBD) (Figure 2a)<sup>15</sup> and  $\alpha$ -helices 4-9 constitute the hormone-binding domain (HBD), with  $\alpha$ -helices 8 and 9 form the homodimerisation interface (Figure 2a). The *apo*-protein and the MmfR-MMF2 complex adopt very similar conformations (Figure 2b).

Ten residues in the hormone-binding site interact directly with MMF2 (Figure 2c). Six form a hydrophobic pocket (L110, A113, W147, L151, V178, and F181) that accommodates the alkyl chain. The carboxylate group of the hormone accepts hydrogen bonds from the Y85 hydroxyl group, the backbone N-H group of Y144 and an ordered water molecule, which also interacts with the backbone N-H group of W147. A second ordered water molecule is hydrogen bonded to the hydroxymethyl group of the hormone and the hydroxyl group of Y144.

### **Cryo-EM structure of MmfR-DNA complex**

Attempts to crystallize MmfR bound to its operator were unsuccessful. We thus employed cryo-electron microscopy (cryo-EM) to elucidate the structure of the MmfR-DNA complex. The protein was complexed with a DNA duplex containing MARE1. Single particles of the complex were observed in cryo-EM movies (Figure 2d). Two-dimensional classification of the particles showed they adopt several different orientations (Figure 2d). Three-dimensional classification and subsequent refinement yielded a 4.2 Å density map containing clearly defined secondary structure elements (Figure 2d and Extended Data Figure 7a-d). Superimposition of the MmfR X-ray structure onto the map indicated that the conformation of the DBD changes upon DNA binding (Extended Data Figure 7). Thus, we performed

molecular dynamics flexible fitting simulations to generate a model of the complex (Figure 2e).

Two MmfR homodimers bind to opposite faces of the DNA duplex (Figure 2e, f). Hill coefficients  $>1$  in the FA measurements indicate positive cooperativity in binding of the homodimers to MARE1 and MARE2 (Extended Data Figure 4a). The obtuse angle between the planes that bisect the monomers in each homodimer is  $140^\circ$  (Figure 2f), consistent with other TFTRs that bind as homodimeric pairs (Extended Data Figure 6d-f)<sup>16,18,19</sup>. As in most other TFTR-DNA complexes,  $\alpha$ -helices 2 and 3 of MmfR, encompassing the helix-turn-helix motif, serve as spacer and recognition helices, respectively<sup>20</sup>. The intra-dimer distance between DBDs showed decreases from 47.6 Å in the protein-hormone complex to 37.6 Å in the protein-DNA complex (Figure 2a, e). The C $_{\alpha}$  atoms of the MmfR HBD in the hormone and DNA-bound states were superimposed to understand conformational changes that cause release of the protein from its operator upon hormone binding (Figure 2g). This revealed an upward shift of the DBD towards the HBD in the protein-hormone complex (Figure 2g), which prevents the helix-turn-helix from binding in the major groove of the DNA duplex.

### **Comparison with other ArpA-like TFTRs**

To develop insight into the molecular basis for hormone recognition and signal transduction from the HBD to the DBD, we aligned the sequence of MmfR with other ArpA-like TFTRs of known ligand specificity (Extended Data Figure 8a). The level of residue conservation was mapped onto the structure of the MmfR-MMF2 complex (Extended Data Figure 8b and Figure 3).

The W147, V178 and F181 residues, which form the sides of the alkyl chain binding pocket in MmfR, are very highly conserved in all other members of the ArpA family (Extended Data Figure 8a). While the three residues at the base of this pocket (L110, A113, L151) are highly conserved in AHFCA-binding TFTRs, they are less well conserved in proteins that bind other hormone classes (Extended Data Figure 8a). This suggests that the alkyl chains common to the four known classes of ligand for ArpA-like TFTRs (Figure 1a) likely all bind in this hydrophobic pocket, with differences in residues at the base of the pocket reflecting differences in the length and/or polarity of the alkyl chain.

The side chain NH<sub>2</sub> group of Q130 in MmfR donates a hydrogen bond to the hydroxyl group of Y85, which is in direct contact with the carboxyl group of the AHFCA (Figure 3). Q130 and Y85 are conserved in AHFCA-binding TFTRs, but not other members of the ArpA family (Extended Data Figure 8a), whereas R128 and L129, positioned opposite Q130 on  $\alpha$ -helix 6, are universally conserved in all ArpA-like TFTRs. The guanidinium group of R128 hydrogen bonds to the backbone carbonyl group of S44 and the side chain of L129 forms a hydrophobic contact with aromatic ring of the universally conserved F42 residue, which like S44 is located on  $\alpha$ -helix 1 (Figure 3). Similarly, the carboxylate group of E132, which is also on the opposite face of  $\alpha$ -helix 6 to Q130 and is very highly conserved in ArpA family members, hydrogen bonds to the backbone N-H and guanidinium groups of Y47 and R128, respectively. The side chains of two other very highly conserved residues, I41, located on the opposite face of  $\alpha$ -helix 1 to F42, and V55, located on the top face of  $\alpha$ -helix 2, also form a hydrophobic contact. Polar contacts between R45 in  $\alpha$ -helix 1 and D54 in  $\alpha$ -helix 2, and D120 at the N-terminus of  $\alpha$ -helix 6 and

E36 in  $\alpha$ -helix 1 (via an ordered water molecule) are also quite highly conserved in ArpA homologues. This network of interactions suggests a plausible mechanism for signal transduction from the HBD to the DBD in MmfR. Binding of the carboxylate group of the hormone to the side chain of Y85 forces the C-terminal end of  $\alpha$ -helix 6 downward and the N-terminal end of  $\alpha$ -helix 4 inward (Figure 2g). This pulls the N-terminal end of  $\alpha$ -helix 1 towards the HBD, repositioning the helix-turn-helix (Figure 2g and Extended Data Figure 1b). It seems likely that other ArpA family members employ a similar signal transduction mechanism. However, Y85 and Q130 are not conserved, reflecting the structural differences between AHFCAs and the other hormone types.

We created Y85F and Q130E mutants of MmfR to probe the role played by these residues in hormone binding and signal transduction. Binding of the mutant proteins to MARE1 and dissociation of the resulting complexes by MMF1 was determined using EMSAs and FA measurements (Extended Data Figure 3d and Supplementary Figure 4). In both cases, the mutant proteins bound tightly to the operator with positive cooperativity, but much higher concentrations of MMF1 were needed to dissociate them from MARE1. These results confirm that Y85 and Q130 play an important role in recognition of the hormone and transmission of the signal from the HBD to the DBD in AHFCA-binding TFTRs.

### **AHFCA structure-activity relationship**

Differences in the ability of MMFs1-5 to dissociate MmfR from its operators indicate that the alkyl chain is one determinant of hormone recognition by AHFCA-binding TFTRs. To further probe structural features that are important for hormone



recognition, a library of AHFCAs was synthesised (Extended Data Table 2 and Extended Data Figure 5b, c; Supplementary Figures 5-34). The dissociation of MmfR from MARE1 by these analogues was assessed using EMSAs (Extended Data Table 2 and Extended Data Figure 3e) and FA measurements enabled EC<sub>50</sub> values to be determined in some cases (Extended Data Table 2 and Supplementary Figure 35).

While moderate changes to the alkyl chain can be tolerated, more extensive changes strongly affect activity. Removal of the hydroxyl group from the hydroxymethyl substituent is tolerated, but other changes are not. It is surprising that the analogue with an altered hydroxymethyl substituent retains activity, because Y144 hydrogen bonds to the hydroxyl group via an ordered water molecule (Figure 2c). To verify that interaction of Y144 with the hydroxyl group is not critical for hormone recognition, we created a Y144F mutant of MmfR. FA measurements indicated that the affinity of this for MARE1 and the EC<sub>50</sub> of MMF1 were similar to the wild type protein (Supplementary Figure 4c). Induction of antibiotic production by the library *in vivo* showed analogous trends to DNA release *in vitro* (Extended Data Table 1).

## Conclusions

Even though A-factor and ArpA were discovered fifty and twenty-five years ago, respectively<sup>2,7</sup>, the molecular mechanisms by which they control gene expression has remained unclear. The biosynthesis of several important medicines is controlled by TFTRs in the ArpA subfamily. A detailed understanding of the

molecular interactions between TFTRs, their operators and the hormones that control them could be exploited to improve the production of such molecules.

Here, we have illuminated how binding of AHFCAs to ArpA-like TFTRs triggers antibiotic production in *S. coelicolor*. The crystal structure of MmfR complexed with MMF2 revealed the molecular basis for hormone recognition by AHFCA-binding TFTRs. In combination with the cryo-EM structure of MmfR bound to MARE1, this crystal structure identifies a conserved signal transduction mechanism in ArpA-like TFTRs (Extended Data Figure 1b). The DNA and ligand binding modes, and the signal transduction mechanism differ markedly from TFTRs that regulate antibiotic resistance gene expression (Extended Data Figure 1a).

Only a handful of tetrameric TFTR-DNA structures have been determined in the last two decades, highlighting that X-ray crystallographic analysis remains challenging in these systems<sup>19</sup>. Our demonstration that such complexes can be structurally characterised using cryo-EM will facilitate molecular understanding of other TFTRs.

## References

1. Barka, E. A. et al. Taxonomy, physiology, and natural products of *Actinobacteria*. *Microbiol. Mol. Biol. Rev.* **80**, 1-43 (2016).
2. Cuthbertson, L. & Nodwell, J. R. The TetR Family of Regulators. *Microbiol. Mol. Biol. Rev.* **77**, 440-75 (2013).
3. Corre, C., Song, L., O'Rourke, S., Chater, K. F. & Challis, G. L. 2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by *Streptomyces coelicolor* genome mining. *Proc. Natl Acad. Sci. USA* **105**, 17510-17515 (2008).
4. Flärdh, K. & Buttner, M. J. *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat. Rev. Microbiol.* **7**, 36-49 (2009).
5. van der Heul, H. U., Bilyk, B. L., McDowall, K. J., Seipke, R. F. & van Wezel, G. P. Regulation of antibiotic production in Actinobacteria: new perspectives from the post-genomic era. *Nat. Prod. Rep.* **35**, 575-604 (2018).
6. Willey, J. M. & Gaskell, A. A. Morphogenetic signalling molecules of the streptomycetes. *Chem. Rev.* **111**, 174-187 (2011).
7. Horinouchi, S. & Beppu, T. Hormonal control by A-factor of morphological development and secondary metabolism in *Streptomyces*. *Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci.* **83**, 277-295 (2007).
8. Kitani, S. et al. Avenolide, a *Streptomyces* hormone controlling antibiotic production in *Streptomyces avermitilis*. *Proc. Natl Acad. Sci. USA* **108**, 16410-16415 (2011).
9. Arakawa, K., Tsuda, N., Taniguchi, A. & Kinashi, H. The butenolide signaling molecules SRB1 and SRB2 induce lankacidin and lankamycin production in *Streptomyces rochei*. *ChemBioChem* **13**, 1447-1457 (2012).
10. Bentley, S. D. et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141-147 (2002).
11. O'Rourke, S. et al. Extracellular signalling, translational control, two repressors and an activator all contribute to the regulation of methylenomycin production in *Streptomyces coelicolor*. *Mol. Microbiol.* **71**, 763-778 (2009).
12. Hinrichs, W. et al. Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance. *Science* **264**, 418-420 (1994).
13. Orth, P., Schnappinger, D., Hillen, W., Saenger, W. & Hinrichs, W. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat. Struct. Mol. Biol.* **7**, 215-219 (2000).

14. Le, T. B. K., Schumacher, M. A., Lawson, D. M., Brennan, R. G. & Buttner, M. J. The crystal structure of the TetR family transcriptional repressor SimR bound to DNA and the role of a flexible N-terminal extension in minor groove binding. *Nucleic Acids Res.* **39**, 9433-9447 (2011).
15. Natsume, R., Ohnishi, Y., Senda, T. & Horinouchi, S. Crystal structure of a  $\gamma$ -butyrolactone autoregulator receptor protein in *Streptomyces coelicolor* A3(2). *J. Mol. Biol.* **336**, 409-419 (2004).
16. Bhukya, H., Bhujbalrao, R., Bitra, A. & Anand, R. Structural and functional basis of transcriptional regulation by TetR family protein CprB from *S. coelicolor* A3(2). *Nucleic Acids Res.* **42**, 10122-10133 (2014).
17. Takano, E. et al. A bacterial hormone (the SCB1) directly controls the expression of a pathway-specific regulatory gene in the cryptic type I polyketide biosynthetic gene cluster of *Streptomyces coelicolor*. *Mol. Microbiol.* **56**, 465-479 (2005).
18. Schumacher, M. A. et al. Structural mechanisms of QacR induction and multidrug recognition. *Science* **294**, 2158-2163 (2001).
19. Schumacher, M. A. et al. Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR. *EMBO J.* **21**, 1210-1218 (2002).
20. Ramos, J. L. et al. The TetR family of transcriptional repressors. *Microbiol. Mol. Biol. Rev.* **69**, 326–356 (2005).

## Figure Legends

**Fig. 1 | Classes of Actinobacterial hormone that induce antibiotic production by binding TFTRs and proposed mechanism for regulation of methylenomycin A biosynthesis by AHFCAs.** **a**, Representative structures of the four hormone classes –  $\gamma$ -butyrolactones (GBLs), 2-alkyl-4-hydroxymethyl-furan-3-carboxylic acids (AHFCAs), 4-alkylbutenolides (ABs) and 2-alkyl-4-hydroxy-3-methylbutenolides (AHMBs) – known to control antibiotic production in Actinobacteria. **b**, Generalised mechanism for induction of antibiotic biosynthesis, involving hormone-mediated de-repression of a transcriptional activator by a TFTR. **c**, Proposed mechanism for regulation of methylenomycin A biosynthesis in *S. coelicolor*. MmfR is a TFTR that represses the *mmfLHP* operon, in addition to *mmfR*, *mmyB* and *mmyY*. The AHFCA concentration steadily increases due to low-level expression of *mmfLHP*. Binding of the AHFCAs to MmfR upregulates *mmfLHP* expression, resulting in a feed forward loop. It also releases repression of the *mmyB* transcriptional activator of the methylenomycin biosynthetic genes.

**Fig. 2 | Structures of *apo*-MmfR, and the MmfR-MMF2 and MmfR-MARE1 complexes determined by X-ray crystallography and cryo-EM.** **a**, X-ray crystal structure of the homodimeric MmfR-MMF2 complex; HBD = hormone-binding domain; DBD = DNA-binding domain. **b**, Overlay of the C $\alpha$  atoms of *apo*-MmfR and the MmfR-MMF2 complex (RMSD 0.447 Å for the 159 C $\alpha$  atoms). **c**, Key residues lining the MmfR hormone-binding pocket, highlighting polar interactions (dashed lines) between MmfR (yellow) and MMF2 (cyan), in two cases mediated by ordered water molecules (purple spheres). All distances are in Å. **d**, Representative cryo-EM micrograph of MmfR-MARE1 complex (top, scale bar: 30 nm) out of 1605 collected in two batches. Representative 2-D class averages of single particles (middle) and modelling of MmfR and the MARE1 DNA duplex into the cryo-EM density map using MDFF. **e**, Overall refined structure of the MmfR-MARE1 complex. Two homodimers of MmfR (monomer units in green and blue) bind to opposite faces of MARE1 (backbone in yellow and bases in blue). The monomer units in one of the MmfR homodimers are labelled mA and mB, whereas in the other homodimer they are labelled mA' and mB'. HTH = DNA-binding helix-turn-helix. The intra-dimer distance between DBDs was measured from the backbone nitrogen atoms of residue G64 in  $\alpha$ -helix 3. **f**, View of the MmfR-MARE1 complex rotated 90° about the y axis. **g**, Overlay of the HBD domain C $\alpha$  atoms for MmfR in complex with the hormone (pink) and MARE1 (green), highlighting the different conformational state adopted by MmfR in the two complexes. The inset shows the axis of  $\alpha$ -helix 1 tilts upwards by 40° in the MmfR-MMF2 complex relative to the MmfR-MARE1 complex, causing the HTH to pull away from the DNA major groove.

**Fig. 3 | Insights into the mechanism of signal transduction in MmfR and other ArpA-like TFTRs.** Highly conserved residues (in yellow) proposed to mediate signal transmission from the HBD to the DBD in TFTRs via a network of hydrogen bonding and hydrophobic interactions. The Y85 and Q130 residues connect the carboxyl group of the ligand (in cyan) to  $\alpha$ -helix 6 in AHFCA-binding TFTRs. Hydrogen bonding interactions and ordered water molecules are represented using dotted lines and purple spheres, respectively. The symbols (¥ and §) used to denote different residue types are defined in the legend for Extended Data Figure 8a.

**Acknowledgements.** C.C. acknowledges support of this work by a University Research Fellowship from The Royal Society (UF090255) and by grants BB/M022765/1 and BB/M017982/1 from the UK Biotechnology and Biological Sciences Research Council (BBSRC). G.L.C. is grateful to the Monash-Warwick Alliance (postdoctoral fellowship to H.B.), the University of Warwick (Chancellor's International Scholarship to S.Z., Warwick Collaborative Postgraduate Research Scholarship to N.M. and Institute of Advanced Study Postdoctoral Research Fellowship to P.K.S.) and the ARC Centre of Excellence in Advanced Molecular Imaging for support of this work. Crystallographic data were collected at beam lines IO3, IO4 and I24 at Diamond Light Source, UK and we acknowledge the support of the beam line scientists. We thank Christopher J. Lupton for assisting with instrument set-up for the FA measurements and Ivan Prokes for assistance with acquiring the NMR data.

**Author contributions.**

C.C. designed the expression vector for MmfR. S.Z., K.M.S., G.L.C. and C.C. designed site-directed mutagenesis experiments. S.Z., P.J.H., G.L.C. and C.C. designed the EMSA experiments. H.B. designed the FA experiments. S.Z., N.M. and G.L.C. designed the synthesis of the AHFCAs, analogues and SCB1. N.M., G.L.C. and C.C. designed the experiments to assess methylenomycin induction *in vivo* using the library of AHFCAs and analogues. H.B., M.J.B., H.V. and G.L.C. designed cryo-EM experiments.

S.Z., H.B., P.J.H., D.R. and C.C. overproduced and purified recombinant MmfR. S.Z. and L.S. performed protein MS analyses. S.Z. and K.M.S. created the MmfR



mutants. S.Z. and N.M. synthesised the library of AHFCAs, analogues and SCB1. S.Z., P.J.H. and C.C. performed EMSAs. H.B. performed FA measurements. N.M. and C.C. investigated the *in vivo* activity of AHFCAs and analogues. D.R., P.K.S. and V.F. prepared samples for protein crystallisation, and D.R. and V.F. collected X-ray diffraction data. H.B. prepared samples for cryo-EM analysis and H.V. collected cryo-EM data.

S.Z., H.B., N.M., P.J.H., G.L.C. and C.C. analysed the results of EMSAs. H.B. and G.L.C. analysed the results of FA measurements. D.R., HB and V.M. processed X-ray diffraction data and solved protein structures. S.Z., D.R., H.B., L.M.A., V.F., G.L.C. and C.C. analysed the X-ray crystal structures. H.B., M.J.B. and H.V. processed cryo-EM data. H.B., M.J.B. and G.L.C. analysed the cryo-EM structure. S.Z., H.B., N.M., L.M.A., G.L.C. and C.C. wrote the manuscript with input from the other authors.

#### **Additional Information:**

**Supplementary Information** is available for this paper.

Correspondence and requests for materials should be addressed to G.L.C. ([G.L.Challis@warwick.ac.uk](mailto:G.L.Challis@warwick.ac.uk)) or to C.C. ([C.Corre@warwick.ac.uk](mailto:C.Corre@warwick.ac.uk)).

Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints).

G.L.C. discloses his role as a co-director of Erebagen Ltd; the other authors declare no competing interests.

**Data availability.** The atomic coordinates have been deposited in the Protein Data Bank for apo and liganded MmfR under the accession numbers 7KY1 and 6SRN

respectively. The three-dimensional cryo-EM map of protein-DNA complex has been deposited in the Electron Microscopy Data Bank (EMDB) under the accession number EMD-20781. This EMDB includes (1) raw half maps (2) B-factor sharpened map and (3) mask used for refinement and sharpening.

## Extended Data Legends

**Extended Data Fig. 1 | Comparison of the mode of DNA and ligand binding, and mechanism of ligand-induced DNA release in TFTRs that regulate the expression of genes for antibiotic resistance and biosynthesis.** For clarity, only  $\alpha$ -helices 1-3 forming the DNA-binding domain (DBD) in both repressor types,  $\alpha$ -helices 4 and 6 in the hormone binding domain (HBD) of the biosynthesis repressors and  $\alpha$ -helices 4, 6, 9 and 10 in the antibiotic binding domain (ABD) of the resistance repressors are shown in the schematics. The direction of movement of the DBD upon ligand binding is highlighted with an arrow. The distance between  $\alpha$ -helices 3 and 3' before and after DNA release and the distance between  $\alpha$ -helix 1 and the ligand after release are given in Å. The structures of the antibiotics/hormones that act as ligands are shown and the name of the TFTR each bind to is shown in parentheses. **a**, TetR and SimR are examples of TFTRs that regulate antibiotic resistance. A single homodimer of these proteins binds the operator and the ligand binding site in the ABD is formed by residues from both subunits. **b**, MmfR as an example of a TFTR that regulates antibiotic biosynthesis. Two homodimers bind the operator and the ligand binding site in the HBD is formed by residues from only a single subunit. PDB entries for DNA form of TetR and SimR are 1QPI and 3QZL respectively and, their liganded entries are 2TRT and 2Y30 respectively.

**Extended Data Fig. 2 | Confirmation of the purity and identity of recombinant MmfR and *in vivo* assay for induction of methylenomycin production in *S. coelicolor*.** **a**, Analysis of purified recombinant His<sub>6</sub>-MmfR; left, SDS-PAGE (MWM:

molecular weight marker); right, mass spectrometry analysis: measured (top) and deconvoluted (bottom) mass spectra of His<sub>6</sub>-MmfR (calculated mass = 27835.5 Da). **b**, *In vivo* assay for induction of methylenomycin production upon addition of increasing amounts of MMF1 signalling molecules to growing mycelia of the MMF non-producing strain *S. coelicolor* W81. Methylenomycin production was detected by growth inhibition of the methylenomycin-sensitive strain *S. coelicolor* M145.

**Extended Data Fig. 3 | EMSA data for release of wild type and mutant MmfR from the *mmyB-mmyY* and *mmfL-mmfR* intergenic regions in the presence of increasing quantities of the MMFs, MMF analogues and SCB1.** **a**, Interaction of MmfR with DNA fragments corresponding to the *mmyB-mmyY* intergenic region (230 bp) in response to increasing amounts of MMFs. **b**, Interaction of MmfR with DNA fragments corresponding to the *mmfL-mmfR* intergenic region (194 bp) in response to increasing amounts of MMFs. **c**, Interaction of MmfR with the DNA fragments corresponding to the *mmfL-mmfR* intergenic region (194 bp) in response to increasing amounts of SCB1. **d**, EMSAs showing that Y85 and Q130 of MmfR play an important role in hormone-induced DNA release. Approximately ten times the quantity of MMF1 is required to release the Y85F and Q130E mutants of MmfR from the *mmfL-mmfR* intergenic region (194 bp) than the wild type protein. **e**, Interaction of MmfR with the DNA fragments corresponding to the *mmfL-mmfR* intergenic region (194 bp) in response to increasing amounts of synthetic MMF analogues. Lane 1: isolated DNA fragments (0.1 pmol); lane 2: DNA fragments mixed with protein (0.1 pmol and 1.8 pmol respectively). For **a**, **b** and **d**, lanes 3 to 9: addition of increasing quantities of MMFs (0.8, 4, 8, 14, 20, 40 and 100 nmol respectively) to the protein-DNA complexes. For **c**, lanes 3 to 7: addition of

increasing quantities of racemic SCB1 (0.8, 8, 20, 40 and 400 nmol respectively) to the DNA/protein complexes. For **e**, lanes 3 to 6: addition of increasing quantities of MMF analogues (0.8, 8, 20 and 100 nmol respectively) to the DNA/protein complexes. At least two independent technical replicates of each EMSA were conducted and in all cases similar results were obtained. For gel source data, see Supplementary Figure 1a-d.

**Extended Data Fig. 4 | Results of FA measurements and EMSAs showing that MmfR binds MAREs in the *mmfL-mmfR* and *mmyB-mmyY* intergenic regions and is released from the MAREs by the MMFs.** **a**, Fluorescence anisotropy plots for binding of MmfR to DNA duplexes containing MARE1 and MARE2. Data points are the mean of three independent technical replicates and error bars represent the standard deviations from the mean. The  $K_d$  (nM) and Hill coefficient ( $\eta$ ) calculated from each data set are shown. The structure of the DNA duplex used in each experiment is shown below the plot. **b**, Confirmation from EMSAs that the five naturally occurring MMFs are able to release MmfR from the *mmfL-mmfR* and *mmyB-mmyY* intergenic regions, and  $EC_{50}$  values calculated from FA measurements for release of MmfR from DNA duplexes containing MARE1 and MARE2.

**Extended Data Fig. 5 | Synthetic routes for MMFs and analogues and SCB1.** **a**, Synthetic route to SCB1. **b**, General synthetic route to MMFs and most analogues. **c**, Synthetic route for MMF analogue 1 lacking a 2-alkyl group.

**Extended Data Fig. 6 | Electron density map for MMF2 bound to MmfR, comparison of the overall fold of the MmfR and CprB monomers and X-ray crystal structures of TFTR-operator complexes.** **a**, SIGMAA-weighted mFo-ΔFc electron density omit map for MMF2 bound to MmfR in mesh representation contoured at the 5σ level. **b**, Overall fold of the MmfR monomer. **c**, Overall fold of the CprB monomer (PDB ID: 4PXI). Both structures are colour-ramped from blue to red from the N- to C-terminus. **d**, X-ray crystal structures of TFTRs that bind as pairs of homodimers in complex with their operators. PDB IDs are as follows: 6EN8 (SaFadR), 6C31 (Rv0078), 4JL3 (Ms6564), 5GPC (FadR), 1JT0 (QacR), 4I6Z (TM1030), 2YVH (CgmR), 4PXI (CprB), 5VL9 (EilR) and 4GCT (SImA). **e**, X-ray crystal structures of TFTRs that bind as single homodimers in complex with their operators. PDB IDs are as follows: 1QPI (TetR), 5DY0 (AmtR), 3LSP (DesT), 5UA1 (KstR), 3ZQL (SimR), 5K7Z (AibR), 3VOK (HrtR) and 5YEJ (BioQ). **f**, Side view of the QacR- and CprB-operator complexes highlighting the obtuse angle between the planes that bisect the monomers in each homodimer.

**Extended Data Fig. 7 | Data quality, overall view of model and cryo-EM map fit for the MmfR-MARE1 complex.** **a**, Relion corrected Fourier shell correlation (FSC) curve of protein-DNA complex map. The inset shows the angular distribution of the particle projections. The length of the projection is a direct measure of the number of assigned particles in each direction. **b**, Model construction by fitting the coordinates for the MmfR-MMF2 complex (pink and blue cartoon) into the cryo-EM density map. Different views of the cryo-EM density maps with the protein-DNA complex modelled into it are shown to the right. **c**, Zoomed-in view of the DBD showing how differently it is oriented in the MmfR-MMF2 complex compared to the

MmfR-MARE1 complex. **d**, Local resolution values projected onto the experimental density map.

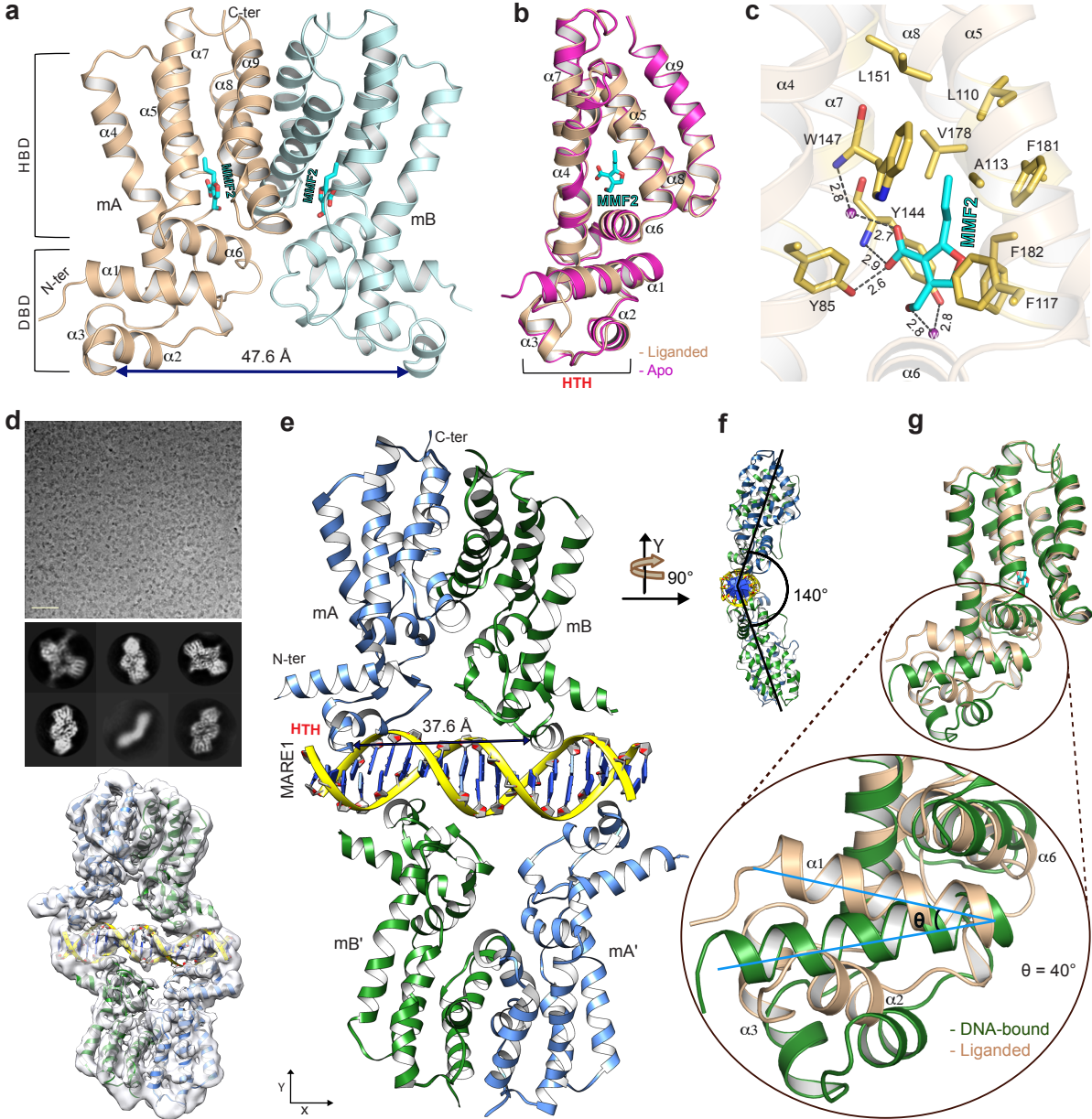
**Extended Data Fig. 8 | Sequence alignment of MmfR with other ArpA-like TFTRs and structural mapping of conserved residues.** **a**, Multiple sequence alignment of TFTRs of known hormone specificity. Amino acids showing a high degree of conservation are coloured yellow, whereas those showing a low degree of conservation are coloured grey. Highly conserved residues hypothesised to be involved in the signal transmission from  $\alpha$ -helices 4 and 6, through  $\alpha$ -helix 1 to  $\alpha$ -helices 2 and 3 in all TFTRs are marked ¥. The highly conserved residues and residue (F182), which is only conserved in AHFCA-binding TFTRs, lining the hydrophobic pocket of the HBD are indicated with \* and £, respectively. The Y85 and Q130 residues, universally conserved in AHFCA-binding TFTRs, are marked §. Protein names are coloured according to the type of ligand each TFTR responds to; cyan: AHFCAs, red: GBLs, purple: AHMBs and blue: ABs. **b**, Mapping of residues showing a high (yellow) and low (grey) degree of conservation onto the structure of the MmfR-MMF2 complex.

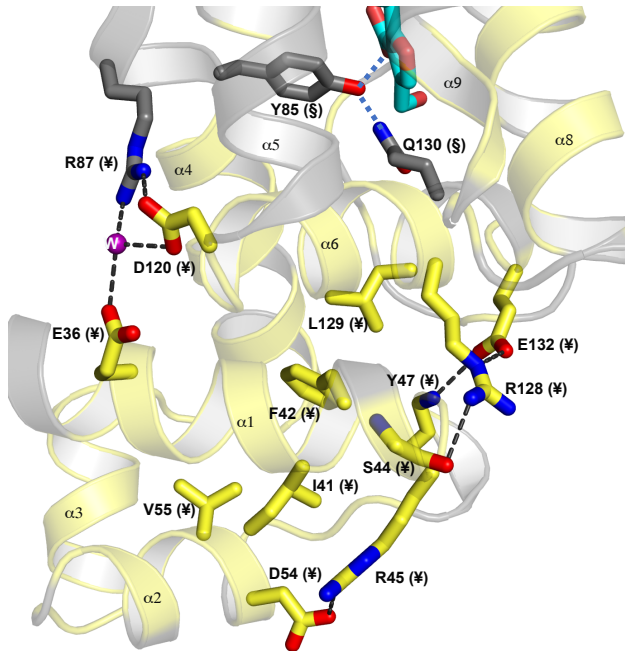
**Extended Data Table 1 | Induction of methylenomycin production by MMFs and analogues, and SCB1 in *S. coelicolor*.** The ability of the five naturally occurring MMFs, the synthetic MMF analogues and SCB1 to induce production of methylenomycin A was compared by observing growth inhibition of methylenomycin-sensitive *S. coelicolor* M145 around agar plugs containing MMF non-producing *S. coelicolor* W81 and five different concentrations of each inducer.

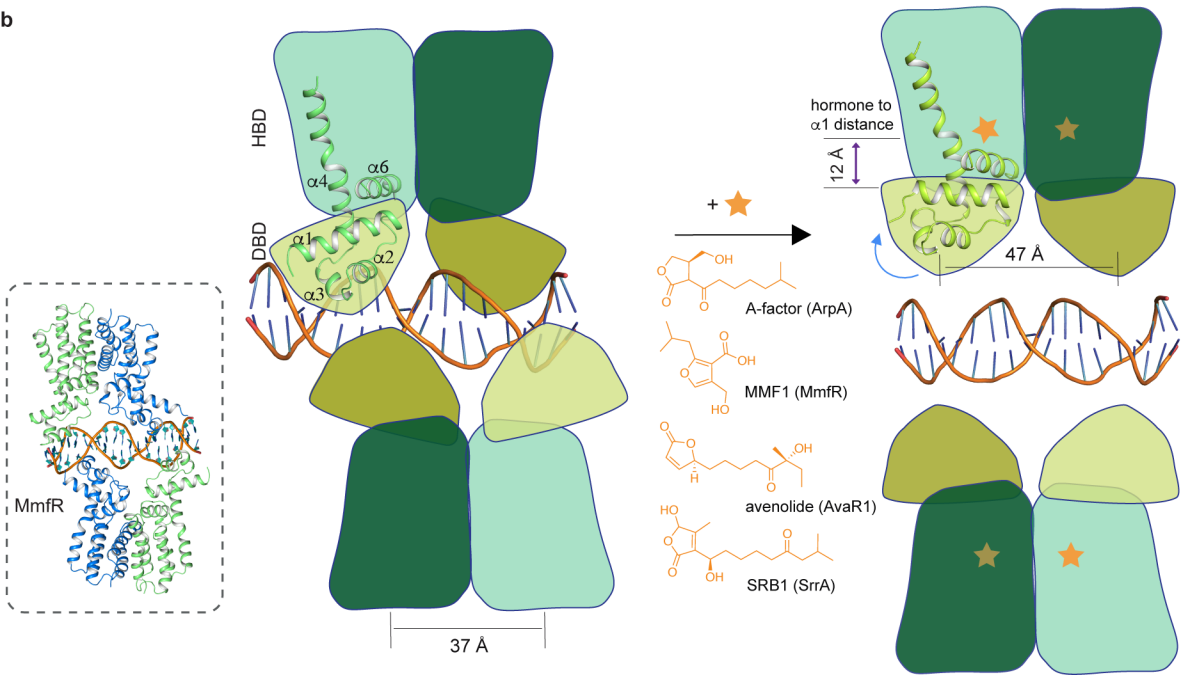
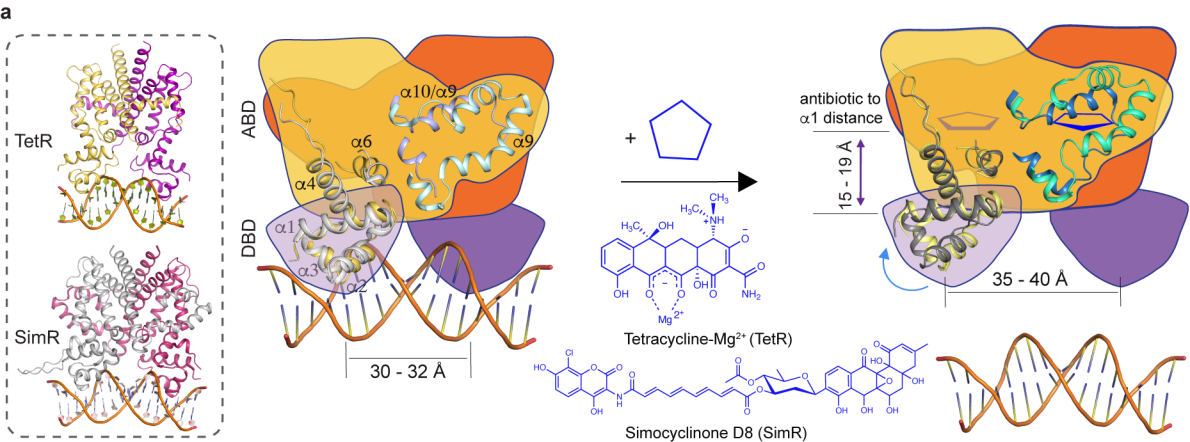
**Extended Data Table 2 | Results of EMSA and FA experiments to probe the role of hormone functional groups in DNA release using an MMF analogue library.** Structural alterations to the alkyl chain included shortening (analogues **1-4**), lengthening (analogues **5** and **6**), desaturation (analogue **7**), altering the position of the methyl branch (analogue **8**) and incorporation of an oxygen atom (analogue **9**). The carboxylic acid group was converted to the corresponding methyl ester (analogue **10**) and the hydroxymethyl group was replaced with a methyl group (analogue **11**) or a hydrogen atom (analogue **12**). Note: **++**, DNA release observed with compound at 8 nmol and above; **+**, DNA release observed with compound at 100 nmol and above; **-**, no release observed; **ND**, not determined.

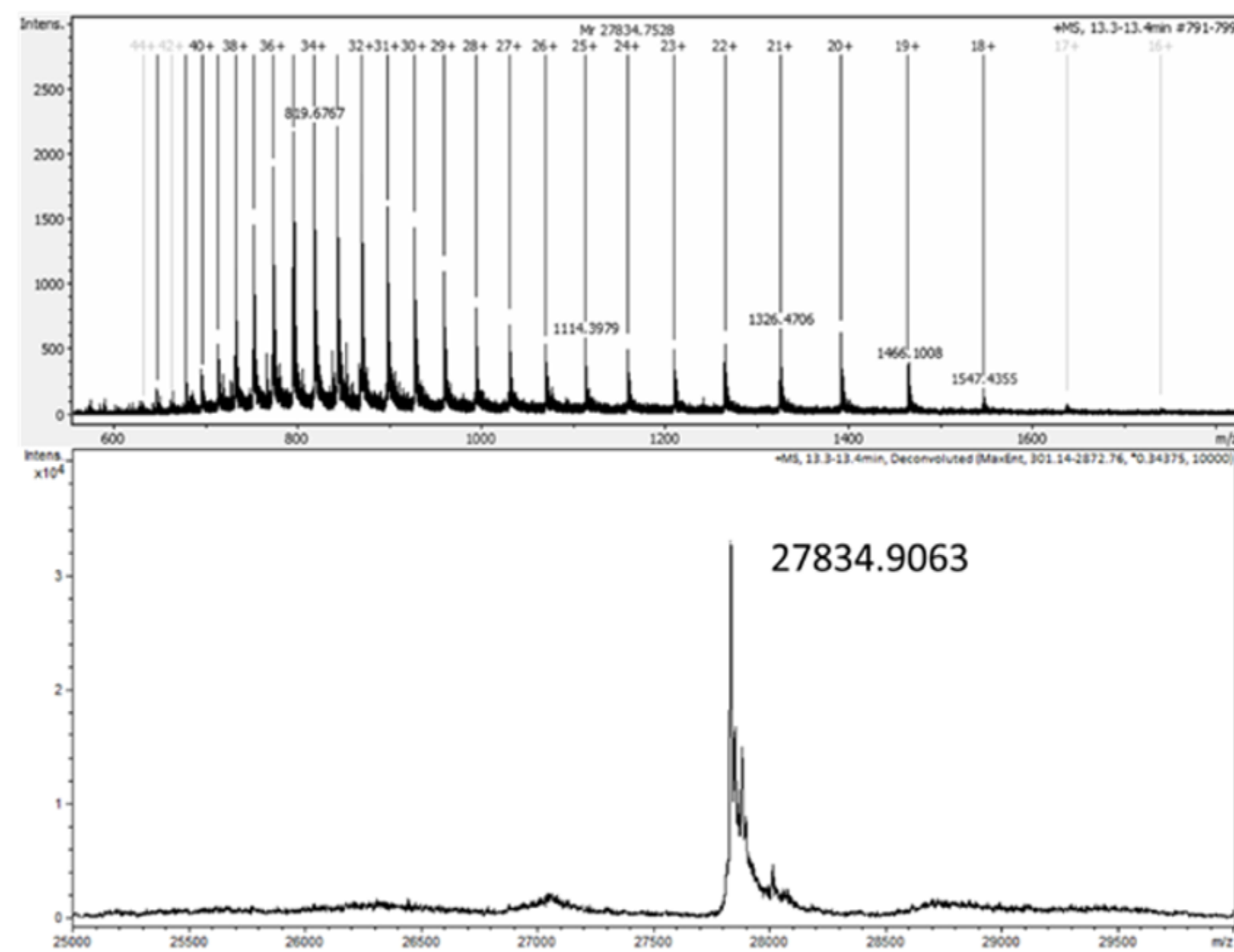
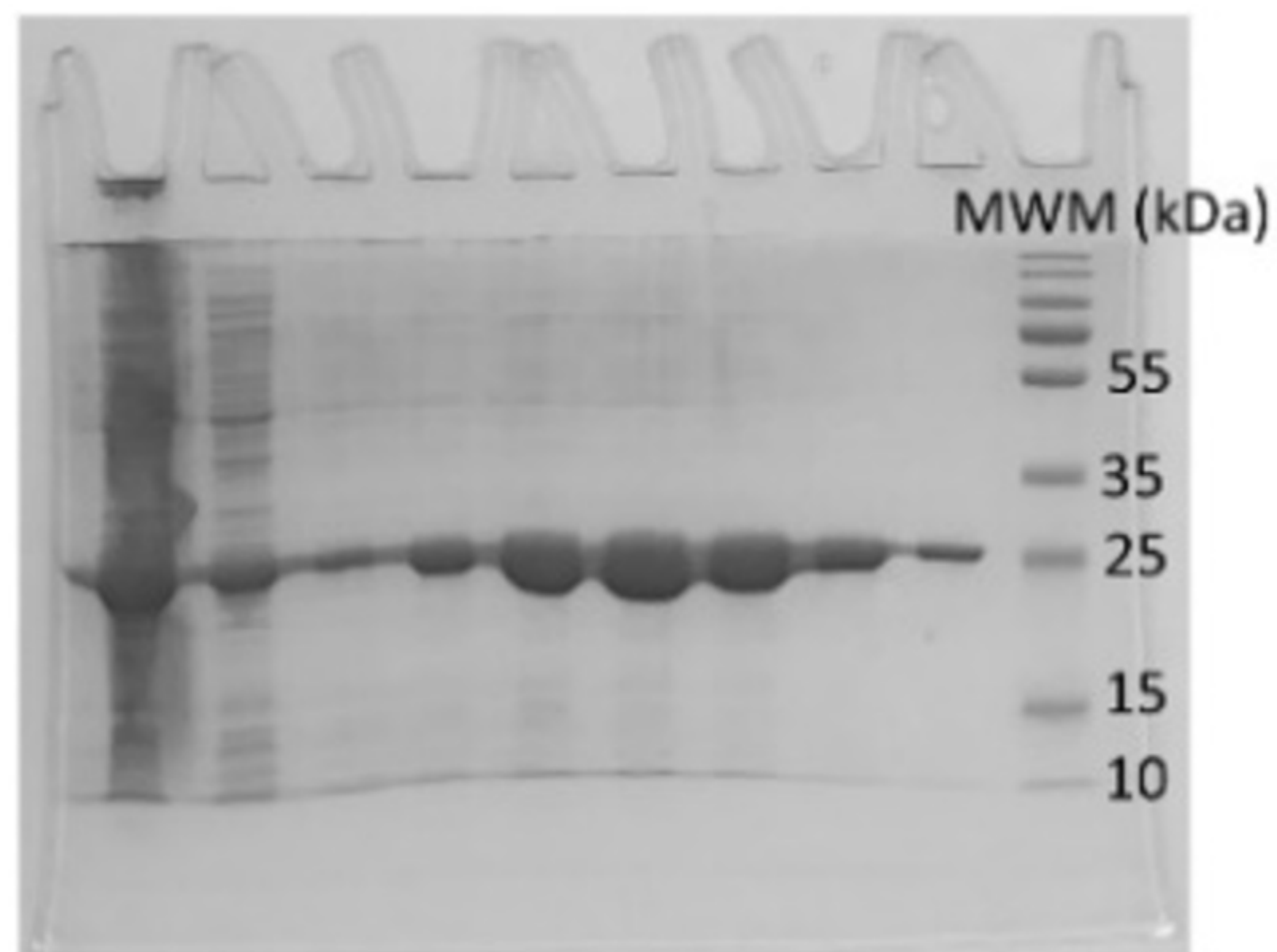
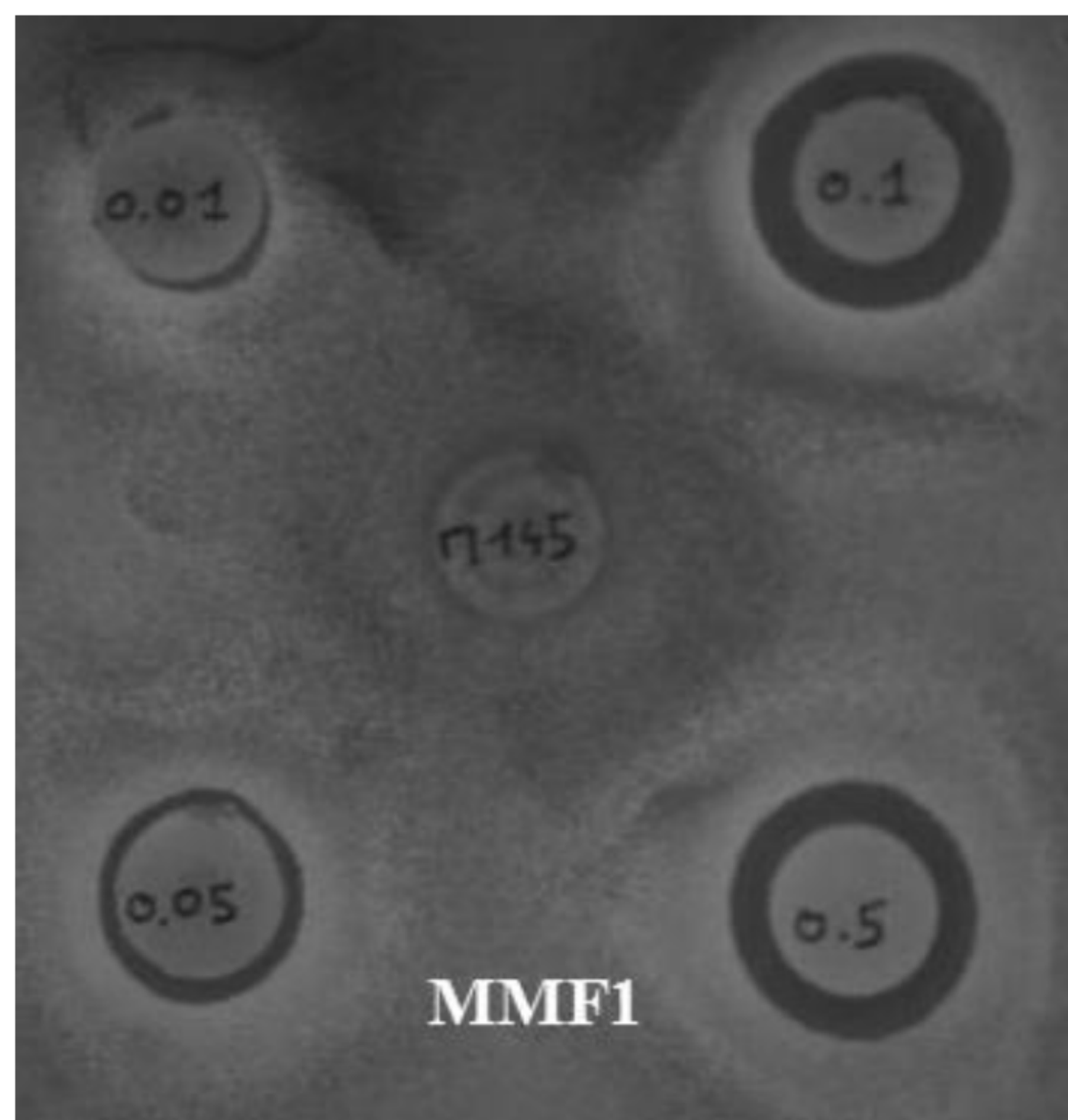




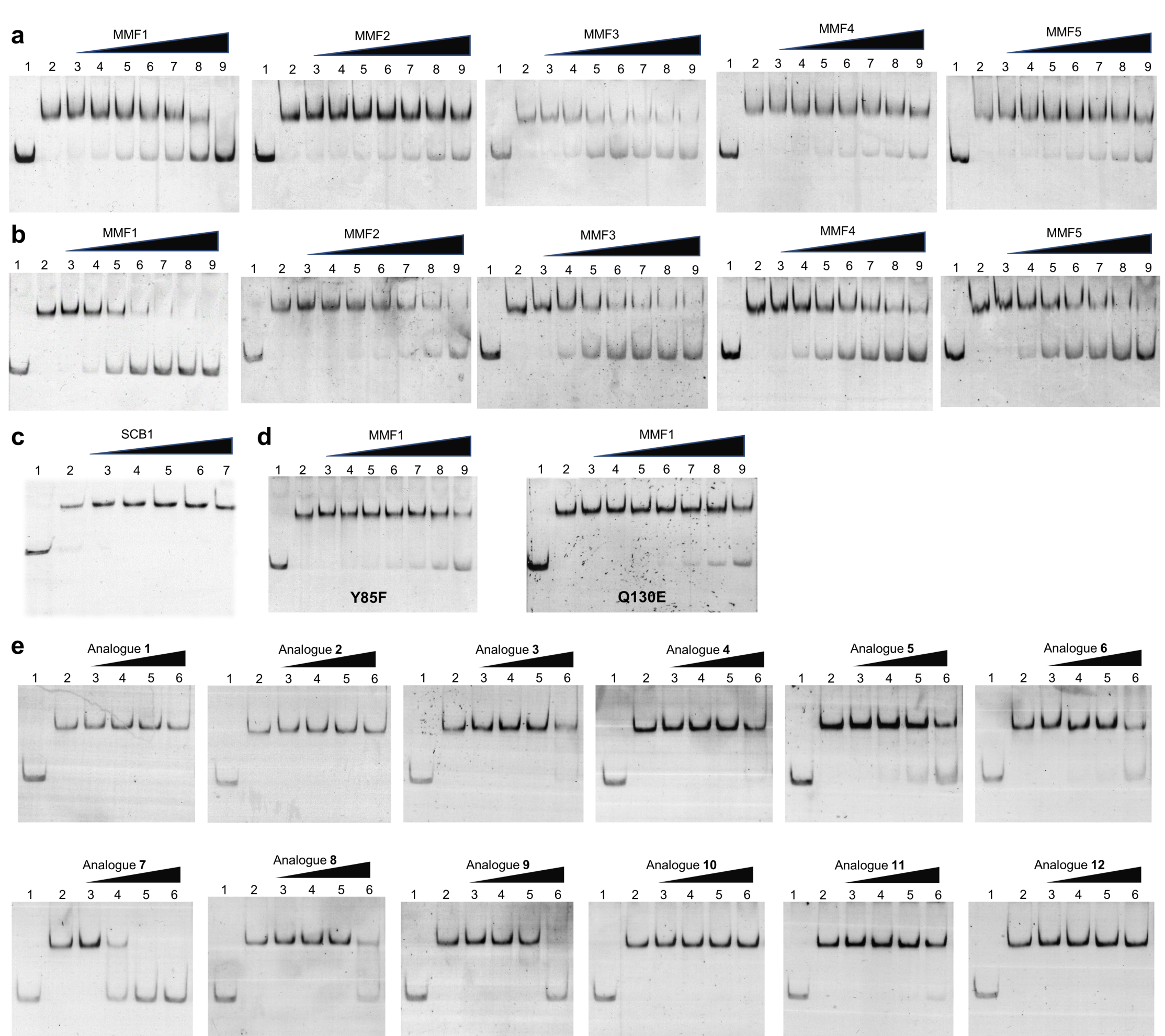




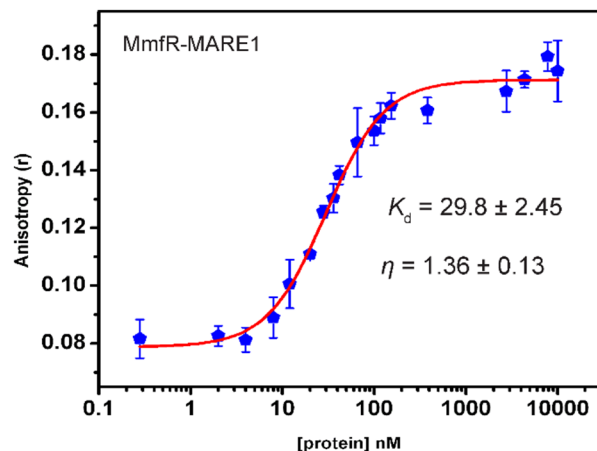


**a****b**



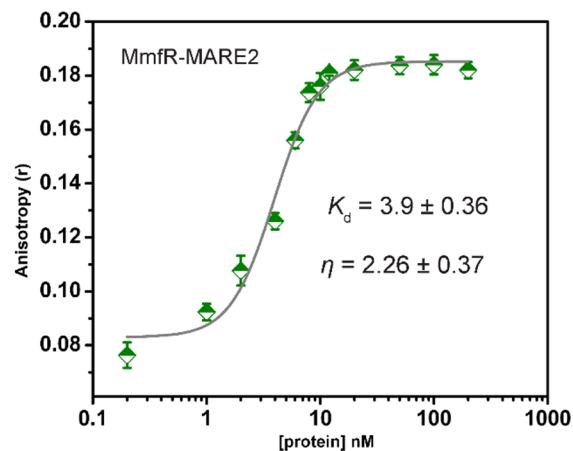


a



A T A T A C C T G C G G G A A G G T A T T A T  
 | | | | | | | | | | | | | | | | | |  
 T A T A T G G A C G C C C T T C C A T A A T A

MARE1  
(*mmfL*-*mmfR*)

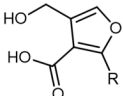


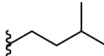
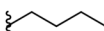
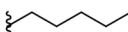
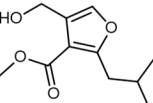




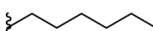
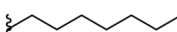
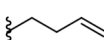

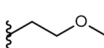
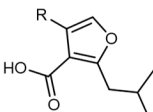


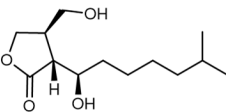


A A A A A C C T T C G G G A A G G T T T G A C  
 | | | | | | | | | | | | | | | | | |  
 T T T T T G G A A G C C C T T C C A A A C T G

MARE2  
(*mmyB*-*mmyY*)

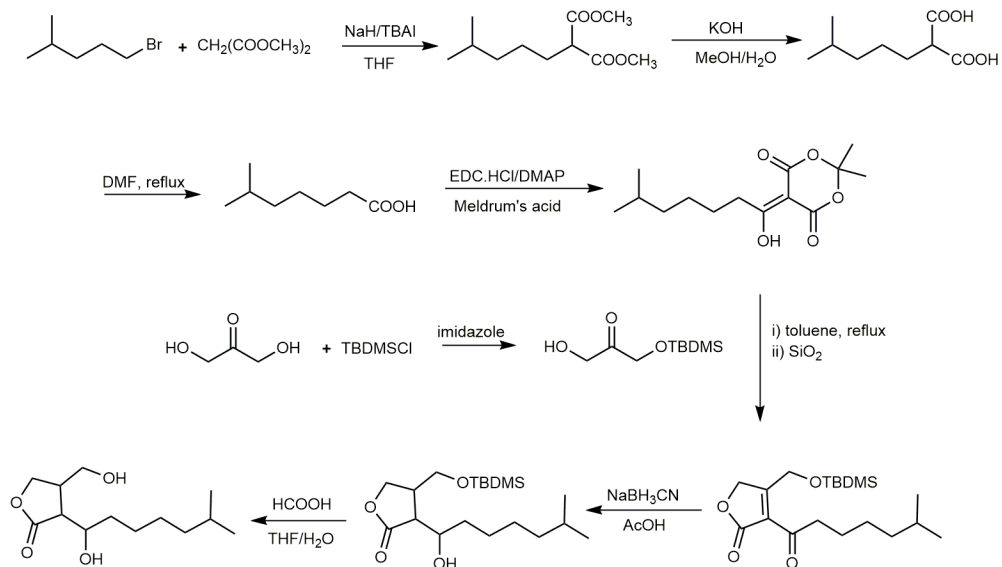
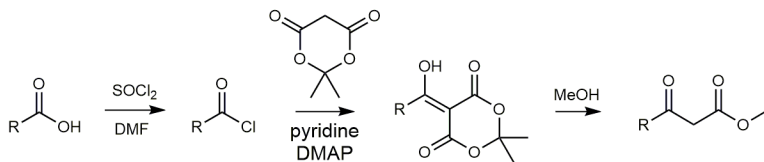
b

Compound	Structure	R=	Release in <i>mmfL</i> - <i>mmfR</i> EMSA	EC <sub>50</sub> (nM) for release from MARE1	Release in <i>mmyB</i> - <i>mmyY</i> EMSA	EC <sub>50</sub> (nM) for release from MARE2
MMF1			yes	0.6±0.04	yes	0.5±0.03
MMF2			yes	3.6±0.21	yes	2.6±0.18
MMF3			yes	1.1±0.05	yes	1.2±0.14
MMF4			yes	1.0±0.08	yes	1.0±0.10
MMF5			yes	0.7±0.06	yes	1.2±0.12

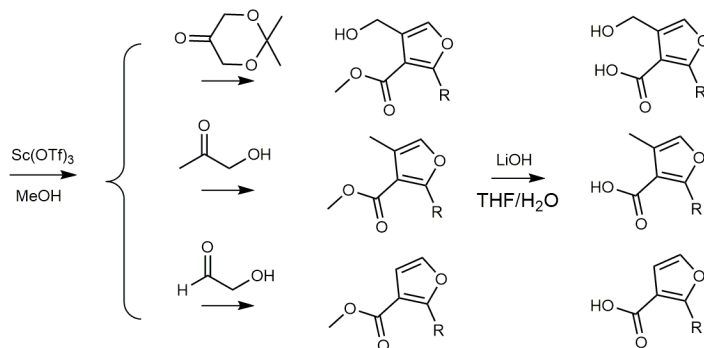
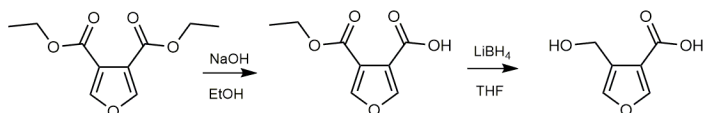
Compound	Structure	R=	Amount (µg)				
			0.05	0.1	0.5	5	50
MMF1			o	o	o	o	o
MMF2			x	x	o	o	o
MMF3			x	x	o	o	o
MMF4			x	o	o	o	o
MMF5			x	o	o	o	o
Analogue 1			x	x	x	x	x
Analogue 2			x	x	x	x	x
Analogue 3			x	x	x	o	o
Analogue 4			x	x	o	o	o
Analogue 5			x	x	o	o	o
Analogue 6			x	x	x	o	o
Analogue 7			x	o	o	o	o
Analogue 8			x	x	o	o	o
Analogue 9			x	x	x	x	o
Analogue 10			x	x	x	x	x
Analogue 11			x	x	o	o	o
Analogue 12			x	x	x	o	ND
SCB1			x	x	x	x	x

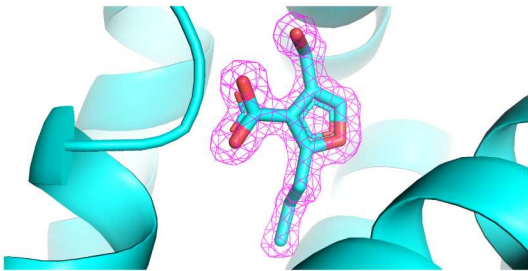
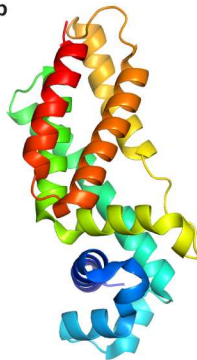
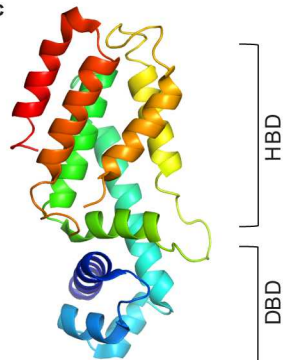
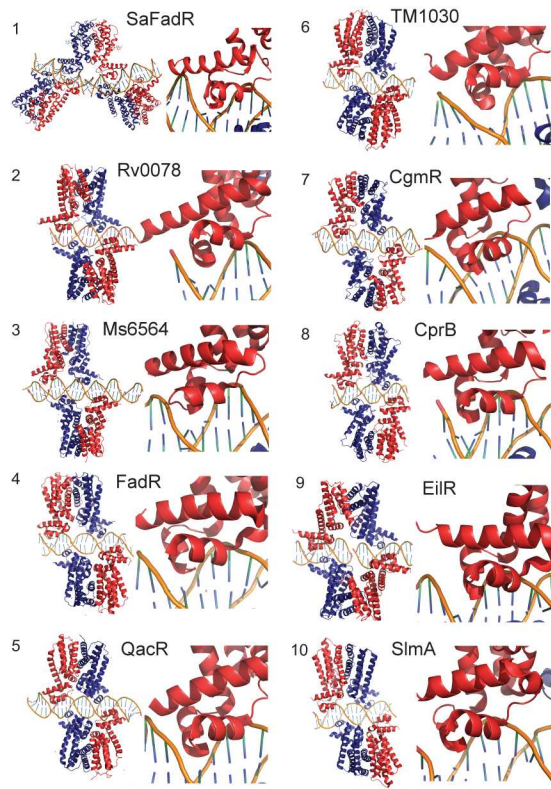
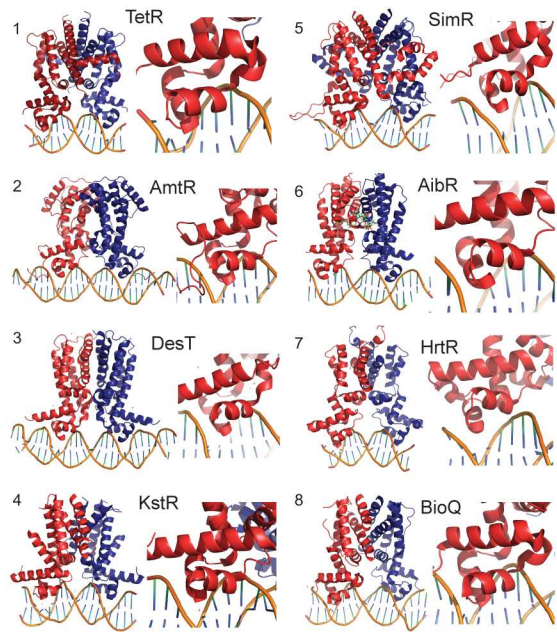
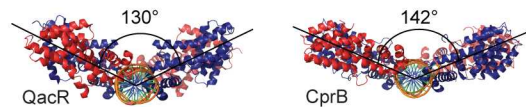
Key: o methylenomycin production was observed; x no methylenomycin production; ND not determined.

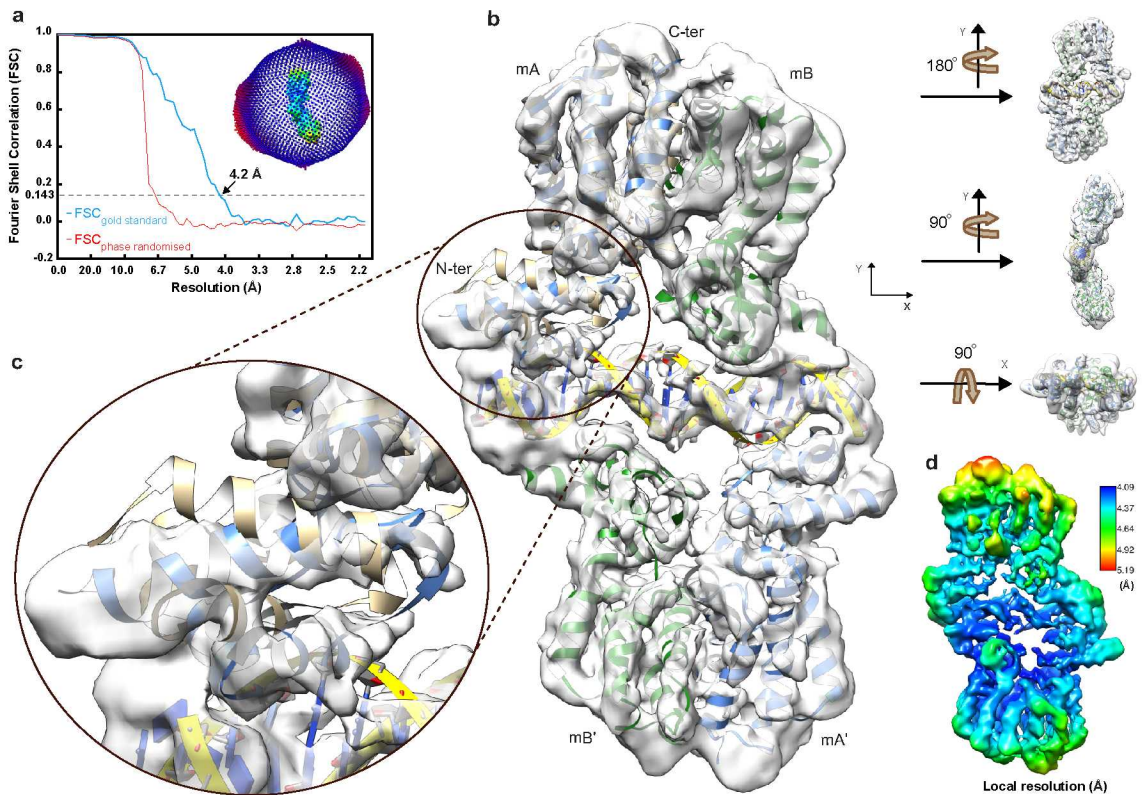


**a****b**

R= alkyl chain

**c**

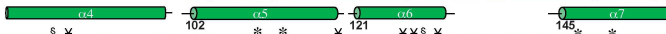
**a****b****c****d****e****f**





Mmfr  
ScIM1  
SgnR  
CprB  
ArpA  
AvaR1  
BarA  
ScbR  
SrrA  
FarA

M TSA-QQPTPF AVRSNVPRGPH PQQERS I KTRAQ ILEAASE IFASRGYRGASVKDVAERVGMTKGAVYFHFPSKE SLAIA  
 MDSKAATHGP-----VAHSRREPKQERARQTKRLVLRATAAEAFKEHGYPVTLQAVAERASMTKGAVYFHYANKEALAVA  
 -----MATPRS QPKQERARRTKVHILQSAAE LFAERGYATVTLQDVAERAEMTKGAVYFHYTNKEALAVA  
 -----MARQLRAEQTRATIIGAAADLFDRRGYESTTLSEIVAHAGVTKGALYFHFPAKEDIHA  
 -----MAKQARAVQTWRSIVDAAASVFDDYGYERAAISEILRRAKVTKGALYFHFASKEAIAQA  
 -----MARQERAI RTRQTILVAAAEVFDEVGYEAAITSDVLKRSVGTKGALYFHFFTSKQELAQA  
 -----MAVRHERVA VRQERAVRTRQAVIRAAASVFDEYGF EAATVAEILSRASVTKGAMYFHFASKEELARG  
 -----MAKQDRAIRTRQ TILDAAAQVFEKQGYQAATITEILKVAGVTKGALYFHFQSK EELALG  
 -----MAQQERAIRTRAVLEAAATVFAEHGYAAATVADILKVAGLTKGALYFHFPSKEALARG  
 -----MAEQVRAIRTRQAILSAARVFD ERGYQAATISEILTVAGVTKGALYFHFQSKEDLAQG



MmfR  
ScIM1  
SgnR  
CprB  
ArpA  
AvaR1  
BarA  
ScbR  
SrrA  
FarA

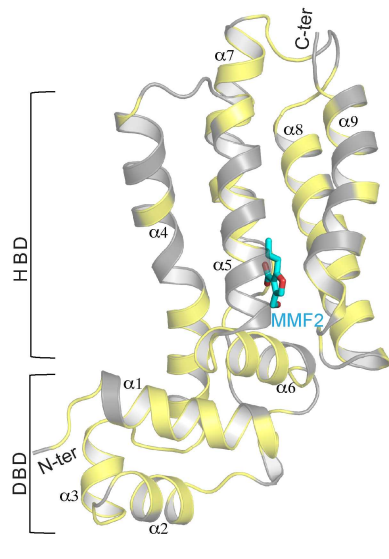
VVEEHYARWPAAMEEIRIQGF<sup>1</sup>PLE<sup>2</sup>TV<sup>3</sup>EEM<sup>4</sup>LHRAAQAFRD<sup>5</sup>D<sup>6</sup>PVMQAGARLQ<sup>7</sup>SERAFT<sup>8</sup>--DAELPLPYVDW<sup>9</sup>THL<sup>10</sup>LEV<sup>11</sup>PLQD<sup>12</sup>  
VIMEHYARWEPLVSEVRSRGLP<sup>13</sup>PLE<sup>14</sup>TLLAV<sup>15</sup>LEGAAEAFRD<sup>16</sup>DVMVQAGARLQ<sup>17</sup>IERSLI<sup>18</sup>--KADLPVPYVWGQ<sup>19</sup>ELLTSLITE<sup>20</sup>  
VVQEHYARWPEILKGAEGDHAEPF<sup>21</sup>DMLTAVLDTV<sup>22</sup>TRAFARD<sup>23</sup>IVVQAGARLQ<sup>24</sup>IERALI<sup>25</sup>--DAELPEPYVGWEDYL<sup>26</sup>TRIAE<sup>27</sup>  
ILEIQSRTSRRLAKDLGRGYSSLE<sup>28</sup>ALMRLT<sup>29</sup>FGMARLCVQGPVLRAGLR<sup>30</sup>LATAGVPVR<sup>31</sup>--PPLPHFTEWREIATSR<sup>32</sup>LLD<sup>33</sup>  
IMDEQTST-VEFEQE-GSP<sup>34</sup>QSLVDGGQQFAFAL-R---HNSMARAGTRL<sup>35</sup>SIEGVF---LGGPHPWGDWD<sup>36</sup>ATARMLEL<sup>37</sup>  
VLAEQVASLPRVPEQ-ELK<sup>38</sup>IQQSLDEALL<sup>39</sup>LAHLL-REGTGDP<sup>40</sup>IVQGSVRLTVDQGS<sup>41</sup>PRDHLNRRVPMQAWTEHTQS<sup>42</sup>LFEE<sup>43</sup>  
VLAEQTLH-VAVPES-GSKAQELVDLT<sup>44</sup>MLVAHGM-L---HDPILRAGTRLALDQ<sup>45</sup>GAV--DFSDANPFGEWGDI<sup>46</sup>CAQLLAE<sup>47</sup>  
VFDAQEPP-QAVPEQ-PLRLQELIDMGM<sup>48</sup>LFC<sup>49</sup>HL-R---TNVVARAGVRLSMDQQA<sup>50</sup>H--GLDRGPFRRWHETLLK<sup>51</sup>LLN<sup>52</sup>  
ILEAQVQP-QLVPQ-QLK<sup>53</sup>LEQWVDAGMTL<sup>54</sup>AHLQ<sup>55</sup>L-P---RDPVLRAGARLSAETHGS--EQHGSAPFTWIAFSAL<sup>56</sup>LEQ<sup>57</sup>  
VLTQ<sup>58</sup>NEDL-L<sup>59</sup>PER-PAK<sup>60</sup>LEQVDAV<sup>61</sup>MLHT<sup>62</sup>HL-R---TNPMV<sup>63</sup>RAGVRLSLDVNA--GGDRSAPFRNW<sup>64</sup>DK<sup>65</sup>FD<sup>66</sup>LLLE<sup>67</sup>

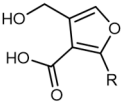





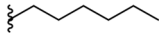
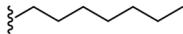
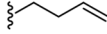

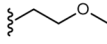
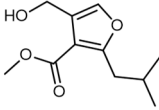
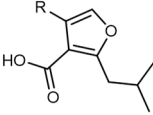




Mmfr  
ScIM1  
SgnR  
CprB  
ArpA  
AvaR1  
BarA  
ScbR  
SrrA  
FarA

AREAGQLRAGVDPAAAARSILVAAFFGMQHVSDNLHQRADIMERWQELRELIMFFALRA-----  
ARDAGQLRADVDPAAALARVVVAFFGSQHISDVLNGRADLMQRHGELLIEAVFRGAMP-----  
ARDAGQLRDGVEPRAAARVLVSAFFGMQHISDVLSGRSLDTEREELRTVLLEGLRR-----  
AVRQSDVHQDIDVSVAHTLVCSVVGTRVVGSTLEPAGREPRRLAEWMYILIRGMVPTRRARYVTTLAAR-----  
QGERGEVFPQIDPMVSAKIIVASFTGTQLVSESSGRADLRGLQVAEMWRHLIPSLIAHPGVIAHKEPGRVDLAAQAREKA  
ARAKGEILPHADV EALAKLVGAFTGVQVLSRIMTGRADLAERVADLYRHLMPSFAMPGILVRDLDFSPER-----  
AQERGEVLPHVNPCKTGDFIVGCFGLQAVSRVTSDRQDLGHRISVMWNHVLPSIVPASMLTWIETGEER-----  
AKENGELLPHVVTDSADLVGTFAGTQVVSQTVSDYQDLEHRYALIKQHLIPAIIVPSVLAALDLSEER-----  
AKRNGELVGHIEPAETAECVLGSFHHGIQLLSQLTNWADIEQRASALFRHVLPAVAVPSVLVRDLDTAPDR-----  
AQAOGELLPHVVPETADTVITGAYGGVSSMSQALTHEODIGORVNALLRHLMPSIAQPSVLASLHLGSEAEVYLEARQ

b



Compound	Structure	R=	Release in <i>mmfL</i> - <i>mmfR</i> EMSA	EC <sub>50</sub> (nM) for release from MARE1
MMF1			++	0.6±0.04
1			-	ND
2			-	ND
3			+	ND
4			+	ND
5			++	4.4±0.51
6			+	ND
7			++	2.4±0.19
8			+	ND
9			+	-
10			-	-
11			+	4.9±0.57
12			-	-

## Methods

### Strains and plasmids

The strains and plasmids used in this study are summarized in Supplementary Table 1.

### Gene cloning

The *mmfR* gene (UniProt ID: Q9JN89) was amplified from cosmid C73\_787 by PCR using the primers listed in Supplementary Table 2. A CACC sequence was introduced at the 5'-end of the forward primer to allow for directional cloning of the blunt-ended PCR products into pET151/D-TOPO, resulting in the fusion of hexahistidine tag to the N-terminus of the recombinant protein. PCR products were purified using the GeneJET Gel Extraction kit (Thermo Scientific) and ligated with the linearized expression vector using the Champion pET151 Directional TOPO Expression kit (Invitrogen) according to the manufacturer's guidelines. One Shot TOP10 chemically competent *E. coli* cells were transformed with the TOPO cloning reaction mixture. Transformants were selected on LB agar plates supplemented with ampicillin (100 µg/mL). Plasmids were purified from ampicillin-resistant colonies using the GeneJET Plasmid Miniprep kit and the integrity of the cloned gene was confirmed by sequencing (GATC Biotech).

### Protein overproduction and purification

For crystallization and EMSAs, recombinant His<sub>6</sub>-MmfR was overproduced in *E. coli* BL21 Star (DE3) cells carrying the pET151-*mmfR* plasmid. Cells were cultured in LB medium containing 100 µg/mL ampicillin at 37 °C to an OD<sub>600</sub> of ~0.6. Expression was induced with 1 mM IPTG, and growth was continued at 20 °C overnight. Cells were harvested by centrifugation, resuspended in binding buffer A (20 mM HEPES, 0.5 M NaCl, 10% glycerol, 10 mM imidazole, pH 8), lysed by sonication, and clarified by centrifugation for 30 min at 4 °C. Cell lysates were passed through a 5 mL chelating sepharose FF column (GE Healthcare) charged with nickel and equilibrated in binding buffer. The column was washed with 100 mL binding buffer A, then 100 mL binding buffer containing 100 mM imidazole, and the protein was eluted in binding buffer containing 500 mM imidazole. Eluted protein was concentrated using a centrifugal concentrator with a 10,000 molecular weight cut-off (Vivaspin), and further purified by gel filtration chromatography on a Superdex S-200 column (GE Healthcare) equilibrated in buffer B (20 mM Tris-HCl, pH 8, 200 mM NaCl). Eluted protein was buffer-exchanged into buffer C (20 mM Tris-HCl, pH 8.0) and concentrated to 14 mg/mL using a centrifugal concentrator, frozen in aliquots of 100 µL in liquid nitrogen, and stored at -80 °C.

The same procedure was used to produce MmfR for cryo-EM analysis, except for the following. IPTG induction was carried out at 18 °C; the protein was purified on a Ni<sup>2+</sup>-nitrilotriacetic acid (Ni-NTA) sepharose resin (GE Healthcare), eluting with a linear gradient of imidazole (100 – 400 mM) in buffer A; and gel filtration was carried out on a HiPrep 16/60 Sephacryl S-200 HR column (GE Healthcare). The hexahistidine tag was removed by adding 1 mg/mL of purified recombinant His<sub>6</sub>-Tobacco Etch Virus protease to 10 mg/mL of MmfR and incubating at 20 °C for 12 h in buffer B containing 0.1 mM EDTA and 0.2 mM DTT<sup>21</sup>. The reaction mixture was passed through Ni-NTA resin equilibrated in buffer B to separate the His<sub>6</sub>-TEV protease from the cleaved MmfR protein. The flow through was further purified by gel filtration on a HiPrep 16/60 Sephacryl S-200 HR column equilibrated in buffer B. The



eluted protein was concentrated to 27 mg/mL using a 10,000 molecular weight cut-off centrifugal concentrator (Vivaspin).

### **Protein crystallisation and X-ray data collection**

Conditions for the crystallization of purified recombinant His<sub>6</sub>-MmfR were screened using a Honeybee crystallization robot. Sitting drops contained 200 nL of 14 mg/mL His<sub>6</sub>-MmfR and 200 nL reservoir solution. Reservoirs contained 75  $\mu$ L of reservoir solution. Numerous hits were obtained and the best crystals grew in a solution containing 20% PEG 3,350 and 0.2 M of various divalent cations. Following optimization in hanging drops containing 1  $\mu$ L protein and 1  $\mu$ L reservoir solution, X-ray diffraction-quality crystals were grown in 10-15% PEG 3,350 and 0.2-0.25 M magnesium formate. Crystals were removed from sitting drops using a nylon loop, soaked briefly in LV cryo oil (MiTeGen), frozen and stored in liquid nitrogen. Crystals diffracted to 1.9 Å in-house using a sealed tube X-ray generator, and to 1.5 Å using synchrotron radiation at Diamond Light Source on beam line I24 (UK).

Purified recombinant His<sub>6</sub>-MmfR (14 mg/mL) was mixed with 1 mM MMF2 and the resulting mixture was subjected to the crystallization conditions described above. These crystals also diffracted to 1.5 Å, using synchrotron radiation at Diamond Light Source on beam line IO4.

Selenomethionine-labelled protein was prepared by overproducing His<sub>6</sub>-MmfR in the methionine auxotroph *E. coli* B834 as described above, except minimal medium containing selenomethionine instead of methionine was used. The resulting protein was purified as described above, except 5 mM 2-mercaptoethanol was included in all purification buffers, and crystallized in several conditions, with the best crystals grown in a solution containing 8% PEG 8,000, 0.1 M magnesium acetate and 0.1 M sodium acetate, at pH 4.5. These crystals diffracted to 2 Å using synchrotron radiation at Diamond Light Source, and a complete dataset was collected at the peak following a wavelength scan at beam line IO3.

### **X-ray data processing, structure determination and refinement**

All X-ray diffraction data were processed using XDS<sup>22</sup>. Further data handling was carried out using the CCP4 software package<sup>23</sup>. The structure was phased using only peak wavelength data from selenomethionine-labelled His<sub>6</sub>-MmfR crystals using a single-wavelength anomalous dispersion (SAD) approach. The SOLVE program located all seven selenated sites in the protein, and RESOLVE fitted 52% of the residues in the resulting electron density<sup>24,25</sup>. At this stage, model building was continued on the 1.5 Å data set from the isomorphous MmfR-MMF2 complex, and the model was further extended automatically by ARP and manually using O<sup>26,27</sup>. The unliganded structure was solved by molecular replacement using Phaser<sup>28</sup>. We employed the standard maximum likelihood restrained refinement with TLS and isotropic B factors for both structures using REFMAC5<sup>29</sup>. Water molecules were added to the atomic models automatically using ARP at the positions of large positive peaks in the difference electron density map, only where the resulting water molecule fell into an appropriate hydrogen bonding environment. The unliganded structure was further refined in Phenix with TLS and isotropic B factors, and maximum likelihood and X-ray/stereochemistry weight restraints, in addition to the default refinement parameters and water molecules were added using phenix.refine<sup>30</sup> and manual refinement done in COOT<sup>31</sup>. The crystallographic asymmetric unit contains one subunit of the polypeptide chain in

both structures. The structures of the dimers were generated from the crystallographic two-fold symmetry operators. The polypeptide could be unambiguously traced from residues 26-214 and 28-214 in the unliganded and MmfR-MMF2 complex structures, respectively. Data collection and refinement statistics are given in Supplementary Table 3.

## **Cryo-Electron Microscopy**

The protein-DNA complex was prepared by adding annealed MARE1 oligonucleotide (5'-ATACCTGCGGGAAGGTATT-3') to a 0.5-fold molar excess of cleaved MmfR and incubating at 20 °C for 1 h. It was purified by gel filtration using the column and protocol described above for cleaved MmfR and concentrated to 17.0 mg/mL. The MmfR-MARE1 complex was diluted to a concentration of 0.1 mg/mL in buffer B and the resulting solution was spotted onto plasma-cleaned QUANTIFOIL Cu EM grids (hole size R1.2/1.3), which were plunge frozen in liquid ethane using an FEI Vitrobot Mark III (blotting time 3.0 s, 100% humidity, -3 mm blotting force, drain time 1 s at 4 °C). Cryo-EM micrographs were collected using Thermo Fisher Scientific Titan Krios™ equipped with a Gatan K2 Summit™ with Quantum-GIF energy filter operated at 300 kV at zero loss mode. A condenser aperture of 50 micron and FEI Voltage phase plate was used in the objective plane. Two Cryo-EM movie datasets (725 and 880 micrographs, each micrograph containing 50 frames) with 10 s exposure time and a total dose of 50 electrons per Å<sup>2</sup> were recorded in counting mode at a nominal magnification of 165,000 (EFTEM mode), which corresponds to a calibrated pixel size of 0.84 Å. Thermo Fisher Scientific EPU automated data collection software was used for data acquisition. Autofocus was set to keep the defocus at 0.5 micron and the phase plate position was advanced every hour.

The acquired movie frames were corrected for beam-induced translational motion and summed using MotionCor2 (version 2.1.10-cuda8.0) and the pixel size was raised to 1.092 Å by binning the dataset<sup>32</sup>. The summed images were saved and subjected to contrast transfer function (CTF) estimation employing GCTF (version 1.06-cuda8)<sup>33</sup>. Summed images with the best CTF fit, phase estimation and a defocus inside the range of 0.5–3.0 micron were selected manually by inspecting the thon rings fit to the theoretical resolution shells and these micrographs were retained for further image processing. This procedure resulted in datasets of 675 and 770 micrographs for MmfR-MARE1 complex. Subsequently, the micrographs were subjected to particle picking using crYOLO1.13<sup>34</sup>. One hundred images were chosen randomly and about 3,000 particles were picked manually to train the auto-picking module, which was then used to auto-pick particles from the CTF corrected and retained summed images of the complex. The trained module was able to pick 181,409 and 292,512 well-centred particles for the complex at a box size of 140 Å (Supplementary Table 4). The particles were then extracted and combined for initial 2-D classification in cryoSPARC<sup>35</sup> (v2) and later a reference-free 2-D classification in Relion3.0<sup>36</sup> to filter out low-quality particles. This yielded a total of 379,033 particles, which were extracted at a box size of 180 pixels and a pixel size of 1.092 Å. The extracted particles were 3-D classified and an initial 3-D model was generated in Relion. From the 2-D and 3-D classes, it appeared that the complex has no symmetry. Thus, C1 symmetry was applied for all the model refinement steps and the particle size was not changed during data reduction and refinement. The particles resulting



from tight masking and model refinement yielded an overall resolution of 4.2 Å with good angular distribution (Extended Data Figure 7).

### Modelling of the protein-DNA complex

The X-ray crystal structure of unliganded MmfR was manually aligned with the cryo-EM density map using Chimera<sup>37</sup>. MmfR was found to adopt four distinct positions in the map. B-form MARE1 DNA was also manually docked into the cryo-EM density map. The models were combined and the fit was optimised using molecular dynamics flexible fitting (MDFF) using namd2/MDFF<sup>38</sup>. Runs were prepared with the VMD<sup>39</sup> graphical user interface, and NAMD<sup>40</sup> was used with the correction map and the CHARMM36<sup>41,42</sup> force field to perform the MDFF. Figures were rendered in Chimera and PyMol (<http://www.pymol.org>).

### Fluorescence anisotropy measurements

One strand from each MARE-containing duplex was 5'-fluorescently labelled with 6-carboxyfluorescein (FAM) and annealed in 5 mM Tris pH 7.5, 15 mM NaCl 0.1 mM EDTA, pH 8.0 (100 µM final concentration) with a complementary unlabelled oligonucleotide by heating to 95 °C for 5 min and slowly cooling to room temperature<sup>43</sup>. The resulting duplexes were used in steady-state FA titrations to assess wild type and mutant MmfR binding and subsequently release by the MMFs and synthetic analogues. Ligands were dissolved in DMSO (200 mM) and the resulting solutions were diluted to 0.1% using buffer B. All titration experiments were carried out at ambient temperature on a Horiba (Jobin Vyon) – FluoroMax-4 spectrofluorometer. FAM-labelled duplexes were excited at wavelength 485 nm and emission was recorded at 520 nm ( $\lambda_{em}$ ). Both excitation and emission slit widths were set at 4 nm. The path length was 1 cm and the signal integration time was set as 0.15 s. Each sample was measured in 500 µL total volume of buffer B. The final DNA concentration was 10 nM for all measurements and the protein concentration was 100 nM for the DNA release measurements. There was no detectable intrinsic fluorescence of buffering components at  $\lambda_{em}$ . Samples were equilibrated for 5 min after addition of titrant before the change in FA was measured. Experiments were performed in triplicate and anisotropy values were calculated using Supplementary equation 1. The apparent  $K_d$  and Hill coefficient ( $\eta$ ) were determined by plotting the mean FA values against log[protein/small molecule] and the resulting binding isotherms were fitted using non-linear regression to Supplementary equation 2 using Origin 7.0. Error bars represent the standard deviations from the mean and the level of confidence in the curve fitting was 97-100%. The G-factor was determined using a solution of free fluorescein.

### Electrophoretic mobility shift assays

The *mmfL-mmfR* (194 bp) and *mmyB-mmyY* (230 bp) intergenic DNA sequences containing the proposed binding sites for MmfR were amplified from cosmid C73\_787 by PCR using the primers shown in Supplementary Table 2. The EMSAs were run on 6% non-denaturing polyacrylamide gels. In each EMSA, lane 1 contained 0.1 pmol of DNA fragment only and other lanes contained 0.1 pmol of DNA and 1.8 pmol of protein. From lane 3, increasing concentrations of compound, which was dissolved in DMSO, were added. The specific amount of each compound used was listed in the legend of the corresponding figure. For each 20 µL reaction, 4 µL of 5× binding buffer (100 mM HEPES, 50 mM

(NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 150 mM KCl, 5 mM EDTA, 5 mM DTT, 1% (w/v) Tween 20, pH 7.6) was used. DNA, protein, 5× binding buffer and distilled water were first mixed and incubated at room temperature for 15 min. After addition of compound solution, the resulting mixture was incubated at room temperature for 10 min, and 5 µL of loading buffer, containing 0.25× TBE, 34% (v/v) glycerol and 0.2% (w/v) bromophenol blue, were added. The total volume of each reaction did not exceed 25 µL. Samples were separated with a pre-run native polyacrylamide gel in 1× TBE buffer at 80 V at 4 °C until the loading dye had migrated to the bottom of the gel. On completion, the gel was incubated in a solution of GelRed (0.005% v/v) in 1× TBE buffer at room temperature for 30 min with agitation, then visualised on a UV transilluminator.

### **Induction of methylenomycin production**

As described previously, we added solutions of each compound at various concentrations directly to round plugs of AlaMM agar (pH 5.0, allowing optimal diffusion of methylenomycin) with 2-day old cultures of *S. coelicolor* W81 (MMF non-producer) growing on them<sup>3</sup>. Each compound was resuspended in methanol at diverse concentrations so that adding a 10 µL volume to the plug resulted in quantities of compounds ranging from 0.01 to 50 µg. The plugs were placed on an agar plate inoculated with *S. coelicolor* M145 (SCP1<sup>-</sup>, SCP2<sup>-</sup>, methylenomycin-sensitive) and, after 96-120 h of incubation at 30 °C, the extent of growth inhibition around the plugs (resulting from methylenomycin production) was recorded.

### **Site-directed mutagenesis**

The Q130E, Y85F and Y144F mutants of His<sub>6</sub>-MmfR were created using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs) according to the manufacturer's instructions. The primers used are listed in Supplementary Table 2. All constructs were sequenced to confirm the presence of the desired mutations.

### **Synthesis of MMFs and their analogues and SCB1**

#### **General experimental details**

Anhydrous toluene, tetrahydrofuran (THF), N,N-dimethylformamide (DMF) and dichloromethane (DCM) were obtained by distillation from calcium hydride under argon and stored over activated 4 Å molecular sieves. All other reagents and solvents were used as supplied (Sigma-Aldrich, VWR or Fisher Scientific). Flash column chromatography was conducted on Sigma-Aldrich silica gel (40-63 µm, 60 Å). Thin layer chromatography (TLC) was carried out on aluminium backed sheets pre-coated with Merck silica gel 60 F<sub>254</sub> and visualised by UV radiation, phosphomolybdic acid or potassium permanganate. The <sup>1</sup>H- and <sup>13</sup>C-NMR spectra were obtained using a Bruker DPX 300, DPX 400 or Avance III HD 500 MHz spectrometer. Chemical shifts (δ) are given in ppm with reference to the residual solvent peak. Coupling constants (J) are rounded to the nearest 0.5 Hertz (Hz). Multiplicities are given as multiplet (m), singlet (s), doublet (d), triplet (t), quartet (q), quintet (quin), sextet (sext), septet (sept) and nonet (non). Low resolution ESI mass spectra were recorded using an Agilent 6130B single quadrupole spectrometer. High resolution mass spectra (HRMS) were recorded on a Bruker MaXis impact mass spectrometer.

## Synthesis of MMFs and analogues

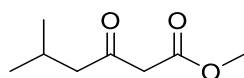
Most of the MMFs and analogues with alteration to the alkyl chains were synthesized according to literature procedures<sup>44,45</sup>. The general synthetic route is shown in Extended Data Figure 5b. It includes preparation of the desired  $\beta$ -ketoester<sup>44</sup>, and subsequent scandium triflate-catalysed condensation with the desired ketone prepared according to published literatures: dihydroxyacetone acetone<sup>46</sup>, hydroxylacetone, hydroxyacetaldehyde, 1-amino-3-hydroxypropan-2-one<sup>47</sup> or mercaptopropanone<sup>48</sup>. MMF analogue **10** is the methyl ester intermediate during MMF1 synthesis. The NMR spectra of **MMFs 1-5** and analogue **10** are identical to the reported data<sup>45</sup>. The MMF analogue **1** without an alkyl chain was produced using a reported method of mono-hydrolysis of the di-ester starting material by dilute reaction condition followed by selective reduction of the ester group in Extended Data Figure 5c<sup>49,50</sup>.

### General procedure for synthesis of $\beta$ -ketoesters

Thionyl chloride (2 eq.) and a catalytic amount of DMF were slowly added to starting acid (1 eq.) under argon. The mixture was stirred overnight at room temperature. The volatile material was removed *in vacuo* using a sodium hydroxide trap. The resulting crude material was used without further purification. To a 0.5 M solution of Meldrum's acid (1 eq.) in dry DCM under argon, dry pyridine (2 eq.) and 4-dimethylaminopyridine (0.2 eq.) were added and the mixture was stirred for 10 min at room temperature before being cooled to 0 °C. Then 1 eq. of the desired acid chloride (or the previous crude mixture) was added dropwise and the reaction was warmed to room temperature and stirred overnight. The resulting mixture was washed with 1 M HCl and water. The organic layer was dried over MgSO<sub>4</sub>, filtered and concentrated *in vacuo*. The resulting dark-yellow oil was diluted in methanol to a final concentration of 0.5 M. The reaction was refluxed for 5 h before the solvent was removed *in vacuo*. The oil residue was purified by flash chromatography (pentane/diethyl ether = 6:1 v/v) to afford the desired  $\beta$ -ketoester as a pale yellow oil.

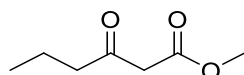
The  $\beta$ -ketoesters for **MMF4** and MMF analogue **2, 3** and **9** are commercially available and used as received.

#### Methyl 5-methyl-3-oxohexanoate



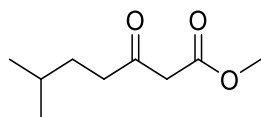
Yield: 60%. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$ : 0.83 (d,  $J$  = 6.5 Hz, 6H), 2.05 (m, 1H), 2.32 (d,  $J$  = 7.0 Hz, 2H), 3.35 (s, 2H), 3.63 (s, 3H). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>)  $\delta$ : 202.3, 167.6, 52.1, 51.7, 49.3, 24.2, 22.3. HR-MS:  $m/z$  calculated for C<sub>8</sub>H<sub>14</sub>O<sub>3</sub> [M+Na]<sup>+</sup>: 181.0835; found: 181.0833.

#### Methyl 3-oxohexanoate



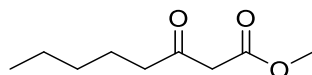
Yield: 55%. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$ : 0.92 (t,  $J$  = 7.5 Hz, 3H), 1.61 (sext,  $J$  = 7.5 Hz, 2H), 2.54 (t,  $J$  = 7.5 Hz, 2H), 3.47 (s, 2H), 3.71 (s, 3H). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>)  $\delta$ : 202.5, 167.5, 51.8, 48.6, 44.4, 16.6, 13.2. HR-MS:  $m/z$  calculated for C<sub>7</sub>H<sub>12</sub>O<sub>3</sub> [M+Na]<sup>+</sup>: 167.0679; found: 167.0683.

*Methyl 6-methyl-3-oxoheptanoate*



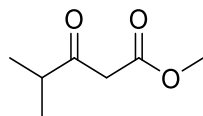
Yield: 21%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.83 (d,  $J = 6.0$  Hz, 6H), 1.54 (m, 3H), 2.54 (m, 2H), 3.48 (s, 2H), 3.73 (s, 3H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 202.9, 179.9, 167.8, 40.9, 33.4, 32.0, 27.4, 22.1. HR-MS:  $m/z$  calculated for  $\text{C}_9\text{H}_{16}\text{O}_3$   $[\text{M}+\text{Na}]^+$ : 195.0992; found: 195.0991.

*Methyl 3-oxooctanoate*



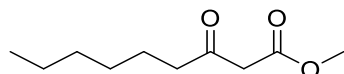
Yield: 80%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.89 (t,  $J = 7.5$  Hz, 3H), 1.29 (m, 4H), 1.57 (br quin,  $J = 7.5$  Hz, 2H), 2.54 (t,  $J = 7.5$  Hz, 2H), 3.46 (s, 2H), 3.69 (s, 3H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 202.3, 167.4, 51.5, 48.4, 42.3, 30.8, 22.7, 22.1, 13.4. HR-MS:  $m/z$  calculated for  $\text{C}_9\text{H}_{16}\text{O}_3$   $[\text{M}+\text{Na}]^+$ : 195.0997; found: 195.0995.

*Methyl 4-methyl-3-oxopentanoate*



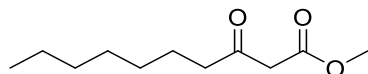
Yield: 38%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 1.04 (d,  $J = 7.0$  Hz, 6H), 2.65 (sept,  $J = 7.0$  Hz, 1H), 3.45 (s, 2H), 3.63 (s, 3H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 206.3, 197.7, 51.9, 46.7, 40.9, 17.6. HR-MS:  $m/z$  calculated for  $\text{C}_7\text{H}_{12}\text{O}_3$   $[\text{M}+\text{Na}]^+$ : 167.0679; found: 167.0979.

*Methyl 3-oxononanoate*



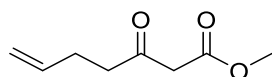
Yield: 30%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.88 (m, 3H), 1.29 (m, 6H), 1.61 (m, 2H), 2.54 (t,  $J = 7.5$  Hz, 2H), 3.46 (s, 2H), 3.72 (s, 3H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 202.8, 167.7, 52.03, 48.7, 42.8, 31.4, 28.5, 23.3, 22.3, 13.8. HR-MS:  $m/z$  calculated for  $\text{C}_{10}\text{H}_{18}\text{O}_3$   $[\text{M}+\text{Na}]^+$ : 209.1148; found: 209.1151.

*Methyl 3-oxodecanoate*



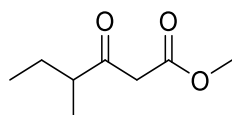
Yield: 85%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.81 (t,  $J = 7.0$  Hz, 3H), 1.22 (m, 8H), 1.53 (m, 2H), 2.47 (t,  $J = 7.5$  Hz, 2H), 3.39 (s, 2H), 3.67 (s, 3H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 202.9, 167.7, 52.2, 49.0, 43.0, 31.6, 29.0, 29.0, 23.5, 22.6, 14.0. HR-MS:  $m/z$  calculated for  $\text{C}_{11}\text{H}_{20}\text{O}_3$   $[\text{M}+\text{Na}]^+$ : 223.1305; found: 223.1304.

#### Methyl 3-oxohept-6-enoate



Yield: 51%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 2.17 (m, 2H), 2.50 (m, 2H), 3.31 (m, 2H), 3.56 (m, 3H), 4.86 (m, 2H), 5.64 (m, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 201.7, 167.4, 136.5, 115.3, 51.9, 48.7, 41.7, 27.2. HR-MS:  $m/z$  calculated for  $\text{C}_8\text{H}_{12}\text{O}_3$   $[\text{M}+\text{Na}]^+$ : 179.0696; found: 179.0696.

#### Methyl 4-methyl-3-oxohexanoate

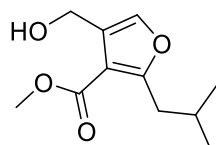


Yield: 34%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.73 (m, 3H), 0.94 (m, 3H), 1.26 (m, 1H), 1.54 (m, 1H), 2.42 (m, 1H), 3.35 (s, 2H), 3.55 (s, 3H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 206.2, 167.7, 51.9, 47.8, 47.2, 25.3, 15.1, 11.1. HR-MS:  $m/z$  calculated for  $\text{C}_8\text{H}_{14}\text{O}_3$   $[\text{M}+\text{Na}]^+$ : 181.0843; found: 181.0843.

#### General procedure for furan cyclisation

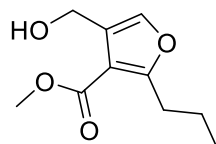
The desired ketone (1 eq.) and  $\beta$ -ketoester (1 eq.) were dissolved in methanol to obtain a 0.5 M solution. Scandium III triflate (0.1 eq.) was added and allowed to stir overnight at room temperature. The solvent was removed *in vacuo* and the residue was purified by flash chromatography (pentane/diethyl ether = 4:1 v/v) giving the desired furan as a pale yellow oil.

#### Methyl 4-(hydroxymethyl)-2-isobutylfuran-3-carboxylate (MMF analogue 10)



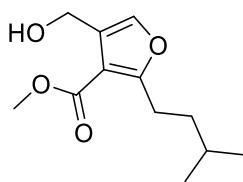
Yield: 76%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.91 (d,  $J$  = 6.5 Hz, 6H), 2.04 (m, 1H), 2.81 (d,  $J$  = 7.0 Hz, 2H), 3.82 (s, 3H), 4.20 (br s, 1H), 4.58 (s, 2H), 7.27 (s, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 165.1, 163.6, 138.3, 126.0, 112.2, 55.8, 51.1, 36.5, 28.1, 22.1. HR-MS:  $m/z$  calculated for  $\text{C}_{11}\text{H}_{16}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 235.0941; found: 235.0941.

#### Methyl 4-(hydroxymethyl)-2-propylfuran-3-carboxylate



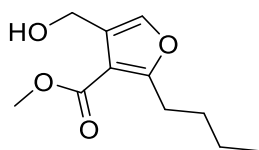
Yield: 84%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.94 (t,  $J$  = 7.5 Hz, 3H), 1.68 (sext,  $J$  = 7.5 Hz, 2H), 2.91 (t,  $J$  = 7.5 Hz, 2H), 3.85 (s, 3H), 3.91 (br s, 1H), 4.56 (s, 2H), 7.25 (s, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 165.4, 164.5, 138.3, 125.9, 111.9, 55.8, 51.4, 29.9, 21.2, 13.6. HR-MS:  $m/z$  calculated for  $\text{C}_{10}\text{H}_{14}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 221.0784; found: 221.0786.

*Methyl 4-(hydroxymethyl)-2-isopentylfuran-3-carboxylate*



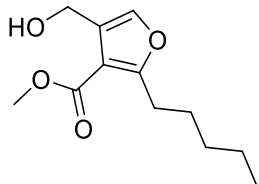
Yield: 83%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.92 (d,  $J$  = 6.0 Hz, 6H), 1.54 (m, 3H), 2.81 (t,  $J$  = 7.5 Hz, 2H), 4.13 (br s, 1H), 3.82 (s, 3H), 4.56 (s, 2H), 7.24 (s, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 165.1, 164.6, 138.1, 126.1, 111.4, 55.8, 51.1, 36.7, 27.5, 25.9, 23.0. HR-MS:  $m/z$  calculated for  $\text{C}_{12}\text{H}_{18}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 249.1097; found: 249.1093.

*Methyl 2-butyl-4-(hydroxymethyl)furan-3-carboxylate*



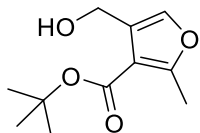
Yield: 80%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.92 (t,  $J$  = 7.5 Hz, 3H), 1.35 (sext,  $J$  = 7.5 Hz, 2H), 1.63 (quin,  $J$  = 7.5 Hz, 2H), 2.94 (t,  $J$  = 7.5 Hz, 2H), 3.85 (s, 3H), 3.91 (br s, 1H), 4.56 (s, 2H), 7.24 (s, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 165.4, 164.7, 138.2, 125.9, 111.8, 55.8, 51.4, 29.9, 27.7, 22.2, 13.6. HR-MS:  $m/z$  calculated for  $\text{C}_{11}\text{H}_{16}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 235.0941; found: 235.0940.

*Methyl 4-(hydroxymethyl)-2-pentylfuran-3-carboxylate*



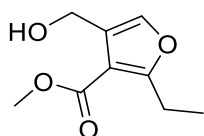
Yield: 76%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.89 (t,  $J$  = 7.0 Hz, 3H), 1.35 (m, 4H), 1.66 (q,  $J$  = 7.5 Hz, 2H), 2.93 (t,  $J$  = 7.5 Hz, 2H), 3.85 (s, 3H), 3.87 (br s, 1H), 4.56 (s, 2H), 7.24 (s, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 165.4, 164.7, 138.2, 125.9, 111.8, 55.8, 51.4, 31.2, 28.0, 27.5, 22.2, 13.8. HR-MS:  $m/z$  calculated for  $\text{C}_{12}\text{H}_{18}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 249.1097; found: 249.1097.

*tert-Butyl 4-(hydroxymethyl)-2-methylfuran-3-carboxylate*



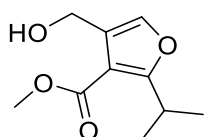
Yield: 68%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 1.47 (s, 9H), 2.40 (s, 3H), 4.01 (br s, 1H), 4.43 (s, 2H), 7.09 (s, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 164.6, 160.0, 137.7, 126.1, 113.9, 81.5, 55.8, 28.2, 14.3. HR-MS:  $m/z$  calculated for  $\text{C}_{11}\text{H}_{16}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 235.0941; found: 235.0940.

*Methyl 2-ethyl-4-(hydroxymethyl)furan-3-carboxylate*



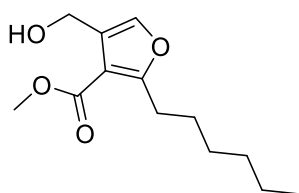
Yield: 85%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 1.22 (t,  $J = 7.5$  Hz, 3H), 2.95 (q,  $J = 7.5$  Hz, 2H), 3.84 (s, 3H), 4.15 (br s, 1H), 4.56 (s, 2H), 7.25 (s, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 165.3, 165.0, 138.2, 126.0, 111.1, 55.8, 51.3, 21.5, 11.9. HR-MS:  $m/z$  calculated for  $\text{C}_9\text{H}_{12}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 207.0628; found: 207.0624.

*Methyl 4-(hydroxymethyl)-2-isopropylfuran-3-carboxylate*



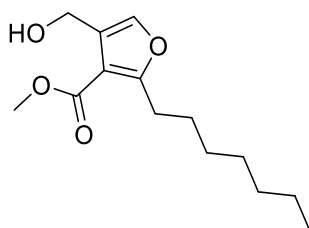
Yield: 79%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 1.25 (d,  $J = 7.0$  Hz, 6H), 3.69 (sept,  $J = 7.0$  Hz, 1H), 3.77 (br s, 1H), 3.86 (s, 3H), 4.56 (s, 2H), 7.25 (s, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 168.5, 165.5, 138.2, 125.8, 110.4, 55.8, 51.6, 27.6, 20.6. HR-MS:  $m/z$  calculated for  $\text{C}_{10}\text{H}_{14}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 221.0784; found: 221.0787.

*Methyl 2-hexyl-4-(hydroxymethyl)furan-3-carboxylate*



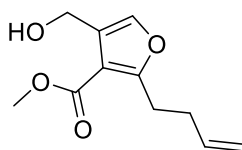
Yield: 40%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.79 (m, 3H), 1.21 (m, 6H), 1.66 (m, 2H), 2.83 (t,  $J = 7.5$  Hz, 2H), 3.76 (s, 3H), 3.86 (br s, 1H), 4.47 (s, 2H), 7.14 (s, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 165.4, 164.6, 138.2, 125.7, 111.8, 55.6, 51.5, 31.4, 28.4, 27.9, 22.5, 13.7. HR-MS:  $m/z$  calculated for  $\text{C}_{13}\text{H}_{20}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 263.1259; found: 263.1260.

*Methyl 2-heptyl-4-(hydroxymethyl)furan-3-carboxylate*



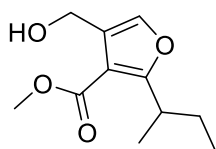
Yield: 70%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.86 (t,  $J = 7.0$  Hz, 3H), 1.27 (m, 8H), 1.62 (m, 2H), 2.90 (t,  $J = 7.5$  Hz, 2H), 3.70 (br s, 1H), 3.85 (s, 3H), 4.53 (s, 2H), 7.22 (s, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 165.8, 165.0, 138.4, 129.9, 112.1, 55.8, 51.8, 31.8, 29.2, 29.0, 28.3, 28.0, 22.7, 14.2. HR-MS:  $m/z$  calculated for  $\text{C}_{14}\text{H}_{22}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 277.1410; found: 277.1408.

*Methyl 2-(but-3-enyl)-4-(hydroxymethyl)furan-3-carboxylate*



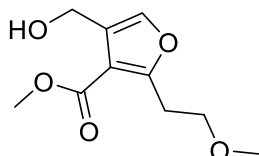
Yield: 82%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 2.41 (q,  $J = 7.5$  Hz, 2H), 3.04 (t,  $J = 7.5$  Hz, 2H), 3.70 (br s, 1H), 3.87 (s, 3H), 4.56 (s, 2H), 4.98 (dd,  $J = 10.5$  and 1.5 Hz, 1H), 5.04 (dd,  $J = 17.0$  and 1.5 Hz, 1H), 5.82 (dddd,  $J = 17.0, 10.0, 6.5$  and 6.5 Hz, 1H), 7.25 (s, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 165.5, 163.7, 138.5, 136.9, 125.9, 115.5, 112.3, 55.8, 51.7, 31.9, 27.8. HR-MS:  $m/z$  calculated for  $\text{C}_{11}\text{H}_{14}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 233.0784; found: 233.0780.

*Methyl 2-sec-butyl-4-(hydroxymethyl)furan-3-carboxylate*



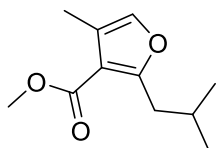
Yield: 68%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.72 (t,  $J = 7.5$  Hz, 3H), 1.12 (d,  $J = 7.5$  Hz, 3H), 1.55 (quin,  $J = 7.0$  Hz, 2H), 3.41 (sext,  $J = 7.0$  Hz, 1H), 3.74 (s, 3H), 3.87 (br s, 1H), 4.47 (s, 2H), 7.17 (s, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 166.7, 164.6, 137.4, 124.9, 110.6, 55.0, 50.5, 33.3, 27.4, 17.5, 10.9. HR-MS:  $m/z$  calculated for  $\text{C}_{11}\text{H}_{16}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 235.0941; found: 235.0944.

*Methyl 4-(hydroxymethyl)-2-(2-methoxyethyl)furan-3-carboxylate*



Yield: 75%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 3.22 (m, 2H), 3.32 (s, 3H), 3.66 (m, 2H), 3.84 (s, 3H), 4.10 (br s, 1H), 4.57 (s, 2H), 7.30 (s, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 164.8, 160.9, 138.8, 126.2, 112.6, 69.8, 58.2, 55.8, 51.3, 28.5. HR-MS:  $m/z$  calculated for  $\text{C}_{10}\text{H}_{14}\text{O}_5$   $[\text{M}+\text{Na}]^+$ : 237.0733; found: 237.0728.

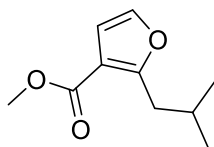
*Methyl 2-isobutyl-4-methylfuran-3-carboxylate*



Yield: 59%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.91 (d,  $J = 6.5$  Hz, 6H), 2.05 (non,  $J = 6.5$  Hz, 1H), 2.13 (s, 3H), 2.82 (d,  $J = 7.0$  Hz, 2H), 3.79 (s, 3H), 7.04 (s, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 164.8, 163.2, 137.7, 120.7, 113.5, 50.9, 36.6, 28.1, 22.1, 9.8. HR-MS:  $m/z$  calculated for  $\text{C}_{11}\text{H}_{16}\text{O}_3$   $[\text{M}+\text{Na}]^+$ : 219.0997; found: 219.0997.



#### Methyl 2-isobutylfuran-3-carboxylate

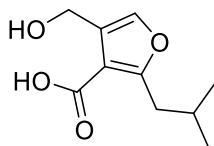


Yield: 75%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.93 (d,  $J = 6.5$  Hz, 6H), 2.08 (non,  $J = 6.5$  Hz, 1H), 2.88 (d,  $J = 7.0$  Hz, 2H), 3.81 (s, 3H), 6.64 (s, 1H), 7.25 (s, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 164.5, 162.7, 140.4, 113.5, 110.5, 51.2, 36.2, 28.3, 22.3. HR-MS:  $m/z$  calculated for  $\text{C}_{10}\text{H}_{14}\text{O}_3$   $[\text{M}+\text{Na}]^+$ : 205.0841; found: 205.0844.

#### General procedure for hydrolysis

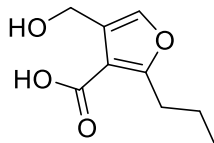
The furan ester (1 eq.) was dissolved in a sufficient amount of 1:1 THF:H<sub>2</sub>O solution. Lithium hydroxide (2.5 eq.) was added and the reaction mixture was stirred for a minimum of overnight at either room temperature or 65 °C. The THF was removed *in vacuo* and the remaining aqueous phase was acidified with concentrated HCl until a white precipitate formed. The precipitate was extracted with diethyl ether and the organic layer was dried over  $\text{Na}_2\text{SO}_4$ , filtered and concentrated *in vacuo* to afford a white or yellowish solid as the desired product.

#### 4-(Hydroxymethyl)-2-isobutylfuran-3-carboxylic acid (MMF1)



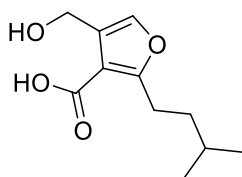
Yield: 85%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.94 (d,  $J = 6.5$  Hz, 6H), 2.09 (m, 1H), 2.88 (d,  $J = 7.0$  Hz, 2H), 4.62 (s, 2H), 7.29 (s, 1H), 8.78 (br s, 2H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 170.0, 166.1, 138.8, 125.3, 112.2, 55.6, 36.8, 28.3, 22.4. HR-MS:  $m/z$  calculated for  $\text{C}_{10}\text{H}_{14}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 221.0784; found: 221.0781.

#### 4-(Hydroxymethyl)-2-propylfuran-3-carboxylic acid (MMF2)



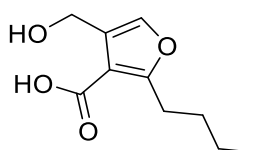
Yield: 88%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.94 (t,  $J = 7.5$  Hz, 3H), 1.71 (sext,  $J = 7.5$  Hz, 2H), 2.97 (t,  $J = 7.5$  Hz, 2H), 4.63 (s, 2H), 7.28 (s, 1H), 8.12 (br s, 2H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 169.8, 166.6, 138.7, 125.3, 111.7, 55.6, 30.0, 21.3, 13.7. HR-MS:  $m/z$  calculated for  $\text{C}_9\text{H}_{12}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 207.0628, found: 207.0624.

*4-(Hydroxymethyl)-2-isopentylfuran-3-carboxylic acid (MMF3)*



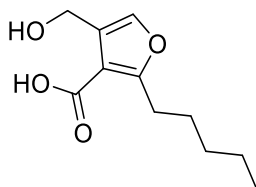
Yield: 97%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.93 (d,  $J$  = 6.5 Hz, 6H), 1.57 (m, 3H), 3.00 (t,  $J$  = 7.5 Hz, 2H), 4.62 (s, 2H), 7.27 (s, 1H), 8.07 (br s, 2H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 169.7, 166.9, 138.7, 125.4, 111.5, 55.6, 36.7, 27.7, 26.0, 22.3. HR-MS:  $m/z$  calculated for  $\text{C}_{11}\text{H}_{16}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 235.0941; found: 235.0942.

*2-Butyl-4-(hydroxymethyl)furan-3-carboxylic acid (MMF4)*



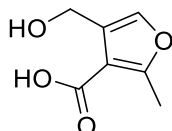
Yield: 77%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.93 (t,  $J$  = 7.5 Hz, 3H), 1.36 (sext,  $J$  = 7.5 Hz, 2H), 1.68 (quin,  $J$  = 7.5 Hz, 2H), 3.00 (t,  $J$  = 7.5 Hz, 2H), 4.62 (s, 2H), 7.28 (s, 1H), 8.07 (br s, 2H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 169.8, 166.8, 138.7, 125.3, 111.6, 55.6, 30.0, 27.9, 22.3, 13.7. HR-MS:  $m/z$  calculated for  $\text{C}_{10}\text{H}_{14}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 221.0784; found: 221.0781.

*4-(Hydroxymethyl)-2-pentylfuran-3-carboxylic acid (MMF5)*



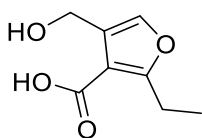
Yield: 93%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.89 (t,  $J$  = 7.0 Hz, 3H), 1.33 (m, 4H), 1.68 (quin,  $J$  = 7.5 Hz, 2H), 2.99 (t,  $J$  = 7.5 Hz, 2H), 4.63 (s, 2H), 7.28 (s, 1H), 8.16 (br s, 2H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 169.7, 166.7, 138.7, 125.3, 111.6, 55.6, 31.4, 28.1, 27.5, 22.3, 13.9. HR-MS:  $m/z$  calculated for  $\text{C}_{11}\text{H}_{16}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 235.0941; found: 235.0941.

*4-(Hydroxymethyl)-2-methylfuran-3-carboxylic acid (MMF analogue 2)*



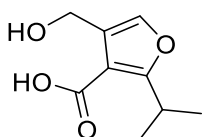
Yield: 72%.  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$ : 2.61 (s, 3H), 4.60 (s, 2H), 6.28 (br s, 2H), 7.27 (s, 1H).  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$ : 170.0, 163.3, 138.8, 125.7, 112.0, 55.8, 14.7. HR-MS:  $m/z$  calculated for  $\text{C}_7\text{H}_8\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 179.0315; found: 179.0306. See Supplementary Figures 5-7.

*2-Ethyl-4-(hydroxymethyl)furan-3-carboxylic acid (MMF analogue 3)*



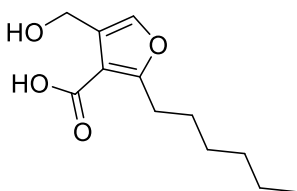
Yield: 93%.  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$ : 1.27 (t,  $J = 7.5$  Hz, 3H), 3.03 (q,  $J = 7.5$  Hz, 2H), 4.60 (s, 2H), 5.70 (br s, 2H), 7.28 (s, 1H).  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$ : 170.0, 168.1, 138.9, 125.7, 111.0, 55.8, 22.0, 12.2. HR-MS:  $m/z$  calculated for  $\text{C}_8\text{H}_{10}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 193.0471; found: 193.0466. See Supplementary Figures 8-10.

*4-(Hydroxymethyl)-2-isopropylfuran-3-carboxylic acid (MMF analogue 4)*



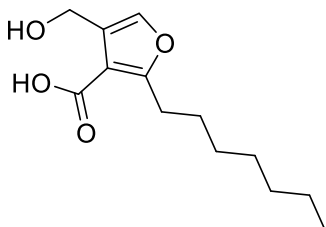
Yield: 95%.  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$ : 1.28 (d,  $J = 7.0$  Hz, 6H), 3.80 (sept,  $J = 7.0$  Hz, 1H), 4.60 (s, 2H), 7.29 (s, 1H).  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$ : 171.1, 170.3, 138.8, 125.5, 110.1, 55.8, 27.8, 20.8. HR-MS:  $m/z$  calculated for  $\text{C}_9\text{H}_{12}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 207.0628; found: 207.0629. See Supplementary Figures 11-13.

*2-Hexyl-4-(hydroxymethyl)furan-3-carboxylic acid (MMF analogue 5)*



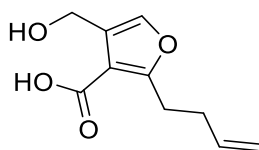
Yield: 76%.  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.89 (t,  $J = 6.5$  Hz, 3H), 1.31 (m, 6H), 1.68 (quin,  $J = 7.5$  Hz, 2H), 3.00 (t,  $J = 7.5$  Hz, 2H), 4.60 (s, 2H), 7.28 (s, 1H).  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$ : 170.3, 167.4, 138.9, 125.6, 111.5, 55.8, 31.6, 29.0, 28.4, 28.0, 22.7, 14.2. HR-MS:  $m/z$  calculated for  $\text{C}_{12}\text{H}_{18}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 249.1097; found: 249.1105. See Supplementary Figures 14-16.

*2-Heptyl-4-(hydroxymethyl)furan-3-carboxylic acid (MMF analogue 6)*



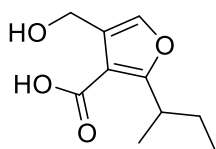
Yield: 86%.  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.88 (t,  $J = 7.0$  Hz, 3H), 1.30 (m, 8H), 1.68 (quin,  $J = 7.5$  Hz, 2H), 2.99 (t,  $J = 7.5$  Hz, 2H), 4.60 (s, 2H), 7.28 (s, 1H).  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$ : 170.3, 167.3, 138.9, 125.6, 111.6, 55.8, 31.8, 29.3, 29.1, 28.4, 28.0, 22.8, 14.2. HR-MS:  $m/z$  calculated for  $\text{C}_{13}\text{H}_{20}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 263.1254; found: 263.1248. See Supplementary Figures 17-19.

*2-(But-3-enyl)-4-(hydroxymethyl)furan-3-carboxylic acid (MMF analogue 7)*



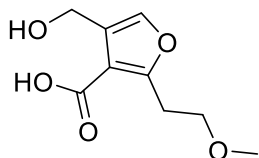
Yield: 97%.  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$ : 2.45 (q,  $J = 7.5$  Hz, 2H), 3.11 (t,  $J = 7.5$  Hz, 2H), 4.61 (s, 2H), 5.00 (dd,  $J = 10.0$  and 1.5 Hz, 1H), 5.06 (dd,  $J = 17.0$  and 1.5 Hz, 1H), 5.84 (ddt,  $J = 17.0$ , 10.0 and 6.5 Hz, 1H), 7.29 (s, 1H).  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$ : 170.1, 166.1, 139.1, 136.9, 125.6, 115.6, 111.9, 55.8, 31.9, 27.9. HR-MS:  $m/z$  calculated for  $\text{C}_{10}\text{H}_{12}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 219.0628; found: 219.0624. See Supplementary Figures 20-22.

*2-sec-Butyl-4-(hydroxymethyl)furan-3-carboxylic acid (MMF analogue 8)*



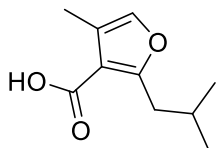
Yield: 87%.  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.85 (t,  $J = 7.5$  Hz, 3H), 1.25 (d,  $J = 7.0$  Hz, 3H), 1.67 (m, 2H), 3.61 (sext,  $J = 7.0$  Hz, 1H), 4.61 (s, 2H), 7.30 (s, 1H).  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$ : 170.4, 170.3, 139.0, 125.3, 111.3, 55.8, 34.5, 28.5, 18.7, 12.1. HR-MS:  $m/z$  calculated for  $\text{C}_{10}\text{H}_{14}\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 221.0784; found: 221.0780. See Supplementary Figures 23-25.

*4-(Hydroxymethyl)-2-(2-methoxyethyl)furan-3-carboxylic acid (MMF analogue 9)*



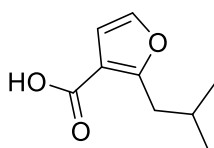
Yield: 95%.  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$ : 3.28 (t,  $J = 6.5$  Hz, 2H), 3.38 (s, 3H), 3.72 (t,  $J = 6.5$  Hz, 2H), 4.59 (s, 2H), 7.31 (s, 1H).  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$ : 168.6, 162.4, 139.5, 125.9, 113.5, 70.3, 58.9, 55.7, 29.0. HR-MS:  $m/z$  calculated for  $\text{C}_9\text{H}_{12}\text{O}_5$   $[\text{M}+\text{Na}]^+$ : 223.0577; found: 223.0577. See Supplementary Figures 26-28.

*2-iso-Butyl-4-methylfuran-3-carboxylic acid (MMF analogue 11)*



Refluxed for 3 days. Yield: 78%.  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.94 (d,  $J = 6.5$  Hz, 6H), 2.10 (non,  $J = 6.5$  Hz, 1H), 2.18 (d,  $J = 1.0$  Hz, 3H), 2.89 (d,  $J = 7.5$  Hz, 2H), 7.08 (d,  $J = 1.0$  Hz, 1H), 12.37 (br s, 1H).  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$ : 171.0, 165.3, 138.2, 121.5, 113.2, 37.0, 28.4, 22.6, 10.2. HR-MS:  $m/z$  calculated for  $\text{C}_{10}\text{H}_{14}\text{O}_3$   $[\text{M}+\text{Na}]^+$ : 205.0835; found: 205.0835. See Supplementary Figures 29-31.

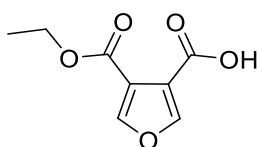
#### 2-iso-Butylfuran-3-carboxylic acid (MMF analogue 12)



Refluxed for 3 days. Yield: 85%.  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$ : 0.95 (d,  $J$  = 6.5 Hz, 6H), 2.11 (non,  $J$  = 6.5 Hz, 1H), 2.91 (d,  $J$  = 7.0 Hz, 2H), 6.70 (d,  $J$  = 2.0 Hz, 1H), 7.28 (d,  $J$  = 2.0 Hz, 1H).  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$ : 170.1, 164.3, 140.9, 113.3, 111.0, 36.5, 28.5, 22.5. HR-MS:  $m/z$  calculated for  $\text{C}_9\text{H}_{12}\text{O}_3$   $[\text{M}+\text{H}]^+$ : 169.0859; found: 169.0860. See Supplementary Figures 32-34.

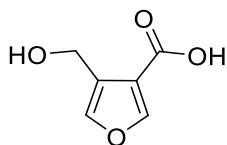
#### Synthesis of MMF analogue 1

##### 4-(Ethoxycarbonyl)furan-3-carboxylic acid



To a solution of diethyl furan-3,4-dicarboxylate (0.1 g, 0.47 mmol) in 2ml ethanol stirred at room temperature was added NaOH (0.018 g, 0.47 mmol) as a fine powder. The reaction was monitored by TLC indicating completion after 3 days. The solvent was removed *in vacuo* and the white solid dissolved in DCM was washed with 5 ml 1M  $\text{NaHCO}_3$  solution. The aqueous layer was acidified with 1M HCl solution and extracted with DCM. The combined organic layers was washed with brine, dried over  $\text{MgSO}_4$  and concentrated *in vacuo*. The residue was purified by flash chromatography (hexane/diethyl ether = 2:1 v/v) to afford 4-(ethoxycarbonyl)furan-3-carboxylic acid (0.044 g, 0.24 mmol, 51%) as white solid.  $^1\text{H}$  NMR (400 MHz, MeOD)  $\delta$ : 1.27 (t,  $J$  = 7.0 Hz, 3H), 4.28 (q,  $J$  = 7.0 Hz, 2H), 8.14 (d,  $J$  = 2.0 Hz, 1H), 8.16 (d,  $J$  = 2.0 Hz, 1H).  $^{13}\text{C}$  NMR (100 MHz, MeOD)  $\delta$ : 166.0, 164.4, 152.3, 151.6, 119.9, 118.3, 63.2, 14.4. HR-MS:  $m/z$  calculated for  $\text{C}_8\text{H}_8\text{O}_5$   $[\text{M}+\text{H}]^+$ : 185.0444; found: 185.0445.

##### 4-(Hydroxymethyl)furan-3-carboxylic acid (MMF analogue 1)

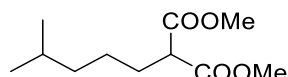


A solution of 4-(ethoxycarbonyl)furan-3-carboxylic acid (0.150 g, 0.82 mmol) in 0.7 ml dry THF was added by syringe to a stirring solution of lithium borohydride (1 M, 0.47 mmol) in THF at 0 °C under argon. The reaction mixture was allowed to warm up to room temperature while being stirred for 3 h. The reaction mixture was then cooled to 0 °C and ice cold THF-water (3 ml, 1:1) was added slowly with vigorous stirring followed by acidification with concentrated HCl. The resulting clear solution was extracted with DCM. The combined organic layers was washed with brine, dried over  $\text{MgSO}_4$  and concentrated *in vacuo* to afford MMF analogue 1 (0.075 g, 0.53 mmol, 64%) as white powder. The NMR spectra of analogue 1 are identical to the reported data<sup>51</sup>.  $^1\text{H}$  NMR (400MHz, MeOD)  $\delta$ : 4.57 (s, 2H), 7.41 (s, 1H), 7.98 (s, 1H).  $^{13}\text{C}$  NMR (100MHz, MeOD)  $\delta$ : 166.9, 150.9, 142.9, 127.2, 119.2, 56.5. HR-MS:  $m/z$  calculated for  $\text{C}_6\text{H}_6\text{O}_4$   $[\text{M}+\text{Na}]^+$ : 165.0164; found: 165.0161.

## Synthesis of SCB1

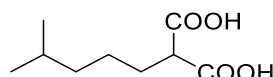
SCB1 was prepared according to modification of a reported procedure for A-factor<sup>52</sup>. The synthetic route for SCB1 is described in Extended Data Figure 5a. The step involving acyl Meldrum's acid was simplified by using EDC/DMAP in one step instead of acid chloride formation and subsequent acylation in two steps. Treatment with sodium cyanoborohydride in acetic acid provided the desired double reduction product of the alkene and ketone in one step, which was followed by deprotection and column chromatography to give racemic SCB1.

### Dimethyl 2-(4-methylpentyl)malonate



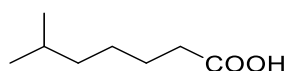
Dimethyl malonate (2.6 ml, 22.1 mmol) was slowly added to a stirring suspension of NaH (0.76 g, 30.0 mmol) in 60 ml THF at 0 °C and the mixture was stirred at the same temperature for 30 min. 1-Bromo-4-methylpentane (3.0 ml, 20.0 mmol) and TBAI (1.51 g, 4.0 mmol) were added and stirred at 0 °C for 30 min before the solution was allowed to warm to room temperature and heated to reflux overnight. The reaction mixture was concentrated *in vacuo*, acidified with 3 M HCl solution and extracted with DCM. The DCM solution was washed with water and brine, dried over anhydrous MgSO<sub>4</sub> and concentrated *in vacuo*. The residue was purified by column chromatography (petroleum ether/ethyl acetate = 10:1 v/v) to give dimethyl 2-(4-methylpentyl)malonate (3.54 g, 16.4 mmol, 82%) as colourless oil. <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ: 0.80 (d, *J* = 6.5 Hz, 6H), 1.09-1.17 (m, 2H), 1.20-1.30 (m, 2H), 1.48 (non, *J* = 6.5 Hz, 1H), 1.82 (q, *J* = 8.0 Hz, 2H), 3.31 (t, *J* = 7.5 Hz, 1H), 3.68 (s, 6H). <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ: 169.8, 52.3, 51.6, 38.3, 28.9, 27.5, 25.0, 22.3. HR-MS: *m/z* calculated for C<sub>11</sub>H<sub>20</sub>O<sub>4</sub> [M+H]<sup>+</sup>: 217.1434; found: 217.1441.

### 2-(4-Methylpentyl)malonic acid



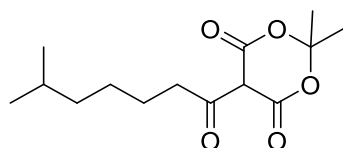
2-(4-Methylpentyl)malonate (3.54 g, 16.4 mmol) was dissolved in a mixture of KOH (9.6 g, 171.4 mmol) in 30 ml water and 60 ml methanol and the mixture was stirred overnight at room temperature. The methanol was removed *in vacuo* and the remaining aqueous solution was washed with DCM, acidified with 3M HCl solution and extracted with DCM. The DCM layers were combined, dried over anhydrous MgSO<sub>4</sub> and evaporated to yield 2-(4-methylpentyl)malonic acid (2.93 g, 15.6 mmol, 95%) as white crystals. <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ: 0.87 (d, *J* = 6.5 Hz, 6H), 1.17-1.27 (m, 2H), 1.33-1.44 (m, 2H), 1.55 (non, *J* = 6.5 Hz, 1H), 1.92 (q, *J* = 8.0 Hz, 2H), 3.44 (t, *J* = 7.5 Hz, 1H), 10.73 (br s, 2H). LR-MS: [M-H]<sup>-</sup> found: 187.1.

### 6-Methylheptanoic acid



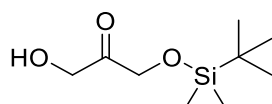
2-(4-Methylpentyl)malonic acid (2.93 g, 15.6 mmol) was dissolved in 30 ml DMF and the mixture was refluxed overnight. 50 ml 5% NaHCO<sub>3</sub> solution was added and the mixture was washed with hexane to remove the neutral impurities. The aqueous solution was acidified with 3 M HCl solution and extracted with DCM. The DCM solution was washed with brine, dried over anhydrous MgSO<sub>4</sub> and concentrated *in vacuo* to give crude product (2.2 g, quantitative) as yellow oil. <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>)  $\delta$ : 0.85 (d, *J* = 7.0 Hz, 6H), 1.14 - 1.21 (m, 2H), 1.27 - 1.38 (m, 2H), 1.52 (non, *J* = 6.5 Hz, 1H), 1.60 (quin, *J* = 7.5 Hz, 2H), 2.34 (t, *J* = 7.5 Hz, 2H), 11.73 (br s, 1H). <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>)  $\delta$ : 180.7, 38.5, 34.1, 27.8, 26.8, 24.9, 22.5. HR-MS: *m/z* calculated for C<sub>8</sub>H<sub>16</sub>O<sub>2</sub> [M+H]<sup>+</sup>: 145.1223; found: 145.1222.

### 2,2-Dimethyl-5-(6-methylheptanoyl)-1,3-dioxane-4,6-dione



EDC·HCl (2.93 g, 15.3 mmol) was added to a solution of 6-methylheptanoic acid (2.20 g, 15.3 mmol) in 20 ml DCM and the mixture was cooled to 0 °C. After stirring for 5 min, 4-dimethylaminopyridine (3.73 g, 30.6 mmol) dissolved in 10 ml DCM was added, followed by Meldrum's acid (2.20 g, 15.3 mmol) dissolved in 10 ml DCM. The reaction mixture was stirred under argon at room temperature for 18 h. The mixture was diluted with 20 ml DCM and washed twice with 1 M HCl, once with brine and once with water. After drying the organic phase over MgSO<sub>4</sub>, the solvent was removed *in vacuo* and the residue was purified by column chromatography to yield acyl Meldrum's acid (2.0 g, 7.4 mmol, 48.4%) as yellow oil. <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>)  $\delta$ : 0.79 (d, *J* = 6.5 Hz, 6H), 1.12-1.17 (m, 2H), 1.28-1.38 (m, 2H), 1.47 (non, *J* = 6.5 Hz, 1H), 1.61 (quin, *J* = 7.5 Hz, 2H), 1.66 (s, 6H), 2.99 (t, *J* = 7.5 Hz, 2H), 15.23 (s, 1H). <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>)  $\delta$ : 170.3, 159.8, 104.4, 91.0, 38.2, 35.4, 27.5, 26.9, 26.4, 26.1, 22.2. LR-MS: [M-H]<sup>-</sup> found: 269.2.

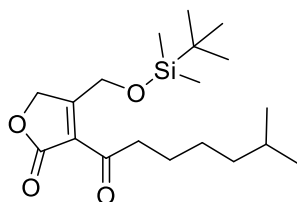
### (1-((*tert*-Butyldimethylsilyl)oxy)-3-hydroxy propan-2-one



*tert*-Butyldimethylsilyl chloride (2.67 g, 17.8 mmol) dissolved in 10 ml DMF was added to a mixture of dihydroxyacetone (5 g, 55.6 mmol) and imidazole (1.51 g, 22.2 mmol) in 50 ml DMF at 0 °C. The reaction was stirred at room temperature for 17 h. 50 ml water was added and the mixture was extracted with diethyl ether. The organic phase was washed with brine, dried over anhydrous MgSO<sub>4</sub> and concentrated *in vacuo*. The residue was purified by column chromatography (petroleum ether/ethyl acetate = 5:1 v/v) to give product (1.82 g, 8.8 mmol, 50%) as colourless oil. <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>)  $\delta$ : 0.12 (s, 6H), 0.90 (s, 9H), 2.99 (br s, 1H), 4.31 (s, 2H), 4.50 (s, 2H), further signals at 3.43-4.24 due to formation of the intermolecular ketal compounds. <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>)  $\delta$ : 210.9, 25.7, 18.2, 5.6,

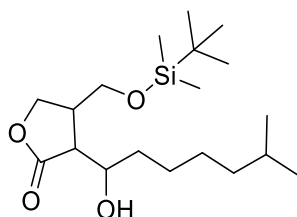
further signals at 63.1-111.4. HR-MS:  $m/z$  calculated for  $C_9H_{20}O_3Si$   $[M+Na]^+$  : 227.1074; found: 227.1076.

*4-(((tert-Butyldimethylsilyl)oxy)methyl)-3-(6-methylheptanoyl) furan-2(5H)-one*



1-((tert-Butyldimethylsilyl)oxy)-3-hydroxy propan-2-one (1 eq.) and acyl Meldrum's acid (1.2 eq.) were added to 2 ml toluene already heated to 110 °C. The reaction mixture was refluxed for 3 h at which point was added 0.5 more equivalents of acyl Meldrum's acid. After stirring for another 5 h, the mixture was allowed to stand at -20 °C for 19 h and purified by column chromatography (hexanes/acetone = 60:1 v/v) without removal of toluene to obtain butenolide as bright yellow oil in 33% yield.  $^1H$  NMR (300 MHz,  $CDCl_3$ )  $\delta$ : 0.09 (s, 6H), 0.84 (d,  $J$  = 6.5 Hz, 6H), 0.89 (s, 9H), 1.10-1.36 (m, 4H), 1.47-1.63 (m, 3H), 2.96 (t,  $J$  = 7.5 Hz, 2H), 4.98 (s, 2H), 5.06 (s, 2H).  $^{13}C$  NMR (75 MHz,  $CDCl_3$ )  $\delta$ : 197.1, 181.2, 170.6, 122.4, 70.2, 61.8, 41.6, 38.7, 29.7, 27.8, 26.9, 25.7, 23.4, 22.6, -5.7. HR-MS:  $m/z$  calculated for  $C_{19}H_{34}O_4Si$   $[M+H]^+$ : 355.2299; found: 355.2306.

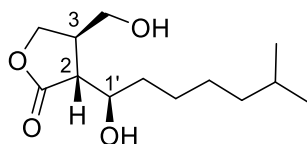
*4-(((tert-Butyldimethylsilyl)oxy)methyl)-3-(1-hydroxy-6-methyl heptyl) dihydrofuran-2(3H)-one*



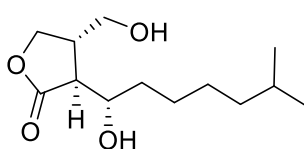
Butenolide (1 eq.) in AcOH was added dropwise to a solution of  $NaBH_3CN$  (2 eq.) in AcOH at 10°C and the mixture was stirred at room temperature for 8 h. After removal of AcOH *in vacuo*, the oily residue was diluted with ethyl acetate and washed three times with 5%  $NaHCO_3$  solution and once with brine. After drying the organic phase over  $MgSO_4$ , the solvent was removed *in vacuo* and the residue was purified by column chromatography (hexanes/ethyl acetate = 7:1 v/v) to yield reduction product as colourless oil in 49% yield. The reduction would result in the formation of three stereogenic centres in the molecule so NMR spectra contain a mixture of different stereoisomers. This resulted in excessive peaks and inaccurate integration. The diastereoisomers were isolated in the next step after deprotection of the silyl group. HR-MS:  $m/z$  calculated for  $C_{19}H_{38}O_4Si$   $[M+Na]^+$ : 381.2432; found: 381.2439.

*(3R,4R)-3-((R)-1-hydroxy-6-methylheptyl)-4-(hydroxymethyl)dihydrofuran-2(3H)-one (SCB1)*





(2*R*,3*R*,1'*R*)-SCB1



SCB1 enantiomer

TBDMS-protected butyrolactone was dissolved in a mixture of THF:HCOOH:H<sub>2</sub>O (6:3:1) and stirred overnight at room temperature. Reactions were brought up to pH 4 with saturated aqueous NaHCO<sub>3</sub> and extracted with ethyl acetate. Ethyl acetate extracts were combined, dried over MgSO<sub>4</sub> and concentrated *in vacuo*. The product was purified by flash column chromatography (hexanes/ethyl acetate = 1:1 *v/v*) to give racemic SCB1 as colourless oil in 43% yield, as well as a pair of their diastereoisomers. The NMR spectra of SCB1 enantiomers are identical to the reported data for SCB1<sup>53</sup>. <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)  $\delta$ : 0.86 (d, *J* = 6.5 Hz, 6H), 1.14-1.20 (m, 2H), 1.26-1.39 (m, 4H), 1.47-1.55 (m, 2H), 1.57-1.64 (m, 1H), 2.65 (dd, *J* = 9.5 and 4.5 Hz, 1H), 2.74-2.81 (m, 1H), 3.68 (dd, *J* = 10.5 and 6.5 Hz, 1H), 3.75 (dd, *J* = 10.5 and 5.0 Hz, 1H), 3.98 (t, *J* = 9.0 Hz, 1H), 4.01 (ddd, *J* = 9.0, 4.5 and 3.0 Hz, 1H), 4.42 (t, *J* = 8.5 Hz, 1H). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>)  $\delta$ : 70.8, 68.3, 63.0, 49.1, 40.1, 38.8, 33.9, 27.9, 27.2, 26.1, 22.6. HR-MS: *m/z* calculated for C<sub>13</sub>H<sub>24</sub>O<sub>4</sub> [M+Na]<sup>+</sup>: 267.1567; found: 267.1565.

**Supplementary Table 1** Strains and plasmids used in this study.

Strains / Plasmids	Relevant properties	Source or Reference
<b>Strains</b>		
<i>Streptomyces coelicolor</i>		
M145	SCP1 <sup>-</sup> , SCP2 <sup>-</sup> (Methylenomycin-sensitive)	[54]
W81	M145 with pCC002:: <i>attB</i> (MMF minus, methylenomycin minus)	[3]
<i>E. coli</i>		
TOP10	One shot chemically competent cells	Invitrogen
BL21 Star (DE3)	F <sup>-</sup> <i>ompT hsdS<sub>B</sub> (r<sub>B</sub><sup>-</sup>, m<sub>B</sub><sup>-</sup>) galdcmrne131</i> (DE3) T7 promoter based expression system	Invitrogen
B834 (DE3)	Met auxotrophic strain for production of Se-Met-MmfR	Novagen
<b>Plasmids</b>		
pET151/D-TOPO	T7 expression vector, ampicillin resistance	Invitrogen
pET151- <i>mmfR</i>	-	This study
pET151- <i>mmfR</i> Q130E	-	This study
pET151- <i>mmfR</i> Y85F	-	This study
pET151- <i>mmfR</i> Y144F	-	This study
C73_787	Integrative cosmid containing the full methylenomycin cluster including the MMFs biosynthetic genes	[55]
pCC002	C73_787 with <i>mmfL</i> → <i>P::apr</i>	[3]

**Supplementary Table 2** Primers used for cloning of *mmfR*, mutagenesis and PCR amplification of the intergenic regions for EMSAs. Fw: forward primer, Rev: reverse primer.

Target	Sequences (5'-3')	Details
<b><i>Protein overproduction</i></b>		
<i>mmfR</i>	Fw: CACCATGACGAGCGCCCAACAAC Rev: TCAGGCGCGGAGAGCGAAG	Cloning of <i>mmfR</i> gene (SCP1.242c – 642 bp) into pET151/D-TOPO vector
<b><i>Site directed mutagenesis</i></b>		
<i>mmfR</i> (Q130E)	Fw: TGCCCGGCTGGAGAGTGAGCG Rev: CCGGCCTGCATCACGGGG	Q5; New England Biolabs
<i>mmfR</i> (Y85F)	Fw: GAGGAGCACTTCGCGCGCTGG Rev: CACCACGGCGATGGCCAG	Q5; New England Biolabs
<i>mmfR</i> (Y144F)	Fw: CCCCTGCCCTTCGTGGACTGG Rev: CAGCTCCGCGTCGATGAAG	Q5; New England Biolabs
<b><i>PCR amplification of intergenic regions</i></b>		
<i>mmfL</i> - <i>mmfR</i> 194 bp	Fw: GGCTGCCTTCCTTCGTGTG Rev: AGGGGCGCTACATCTCCCG	Entire intergenic region between <i>mmfL</i> and <i>mmfR</i>
<i>mmyB</i> - <i>mmyY</i> 230 bp	Fw: GGTGAACTCCTTCGGCGAG Rev: GGCGCCTCACAGTGTCAAAC	Entire intergenic region between <i>mmyB</i> and <i>mmyY</i>

**Supplementary Table 3** Crystallographic data collection and refinement statistics.

	MmfR-MMF2 complex	MmfR (SeMet)	MmfR (Native)
<b>Data collection</b>			
PDB ID	6SRN		7KY1
Space group	P4 <sub>1</sub> 2 <sub>1</sub> 2	P4 <sub>1</sub> 2 <sub>1</sub> 2	C222
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	67.21, 67.21, 92.82	67.04, 67.04, 93.91	60.50, 115.72, 53.09
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength	0.9795	0.9793	0.9686
Resolution (Å)	42.3-1.5 (1.58-1.5) *	93.9-2.2 (2.32-2.2)	53.6-1.50 (1.56-1.50)
<i>R</i> <sub>sym</sub> or <i>R</i> <sub>merge</sub>	0.082 (1.189)	0.137 (0.768)	0.053 (0.659)
<i>I</i> / $\sigma$ <i>I</i>	17.9 (2.2)	18.5 (4.5)	16.0 (2.2)
Completeness (%)	99.9 (99.6)	100 (100)	99.0 (99.5)
Redundancy	11.7 (11.7)	12.7 (12.8)	4.3 (4.1)
CC <sub>1/2</sub>	99.9 (76.2)	99.7 (81.4)	100.0 (84.7)
<b>Refinement</b>			
Resolution (Å)	42.3-1.5		53.6-1.50
No. reflections	34705		29976
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.174 / 0.207		0.189 / 0.213
No. atoms			
Protein	1480		1507
Ligand	13		
Water/Glycerol	256/6		229
<i>B</i> -factors			
Protein	13.4		24.1
Ligand/ion	13.2		
Water	33.4		37.0
R.m.s deviations			
Bond lengths (Å)	0.014		0.007
Bond angles (°)	1.9		0.9
Ramachandran statistics			
Favoured (No., %)	183, 99		189, 98
Allowed (No., %)	2, 1		3, 2
Outliers (No., %)	0, 0		0, 0

**Supplementary Table 4** Cryo-EM data collection and processing for the MmfR-MARE1 complex.

	(Data set 1)	(Data set 2)
<b>Data collection and processing</b>		
Magnification	165,000	165,000
Voltage (kV)	300	300
Electron exposure (e-/Å <sup>2</sup> )	50	50
Defocus range (µm)	0.5	0.5
Pixel size (Å)	0.84	0.84
Initial particle images (no.)	181,409	292,512
Final particle images (no.)		311088
Symmetry imposed		C1
Map resolution (Å)		4.5
FSC threshold		0.143
Map sharpening <i>B</i> factor (Å <sup>2</sup> )		-94.02
Initial model used (PDB code)		4PXI
Model composition		
Non-hydrogen atoms		6777
Protein residues		756
DNA residues		38
Ramachandran plot		
Favored (%)		97.86
Allowed (%)		2.14
Disallowed (%)		0.0
EMDB		EMD-20781

### Supplementary equation 1

$$Anisotropy(r) = \frac{I_{VV} - G(\lambda_{em})I_{VH}}{I_{VV} + 2G(\lambda_{em})I_{VH}}$$

$I_{VV}$  is intensity with excitation and emission polarizers mounted vertically, whereas  $I_{VH}$  is intensity with vertical excitation and horizontal emission polarizers.  $G$  is the so-called  $G$ -factor, which accounts for optical and systematic errors.  $\lambda_{em}$  is the emission wavelength.

### Supplementary equation 2

$$y = r_0 + (r_{max} - r_0) \frac{x^\eta}{k^\eta + x^\eta} \quad (\text{where } r_{max} - r_0 > 0 \text{ for binding and for release it is } < 0)$$

$k$  is the dissociation constant and  $\eta$  is the Hill coefficient. Modified Hill equation.

## References:

21. Nallamsetty, S. et al. Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease *in vivo* and *in vitro*. *Protein Expr. Purif.* **38**, 108-115 (2004).
22. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125-132 (2010).
23. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-763 (1994).
24. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849-861 (1999).
25. Terwilliger, T. C. Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr. D* **59**, 38-44 (2003).
26. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458-463 (1999).
27. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110-119 (1991).
28. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658-674 (2007).
29. Murshudov, G. N. et al. *REFMAC5* for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355-367 (2011).
30. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352-367 (2012).
31. Emsley, P. et al. Features and development of Coot. *Acta Crystallogr. D* **66**, 486-501 (2010).
32. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331-332 (2017).
33. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1-12 (2016).
34. Wagner, T. et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).
35. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290-296 (2017).
36. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
37. Pettersen, E. F. et al. UCFS Chimera-A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612 (2004).
38. Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673-683 (2008).
39. Humphrey, W., Dalke, A. & Schulten, K. VMD-Visual Molecular Dynamics, *J. Mol. Graph.* **14**, 33-38 (1996).
40. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781-1802 (2005).

41. Best, R. B. et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone  $\phi$ ,  $\psi$  and side-chain  $\chi(1)$  and  $\chi(2)$  dihedral angles. *J. Chem. Theory Comput.* **8**, 3257-3273 (2012).
42. Hart, K. et al. Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium *J. Chem. Theory Comput.* **8**, 348-362 (2012).
43. Bhukya, H., Jana, A. K., Sengupta, N. & Anand, R. Structural and Dynamics Studies of the TetR Family Protein, CprB from *Streptomyces coelicolor* in complex with its Biological Operator Sequence. *J. Struct. Biol.* **198**, 134-146 (2017).
44. Oikawa, Y., Sugano, K. & Yonemitsu, O. Meldrum's acid in organic synthesis. 2. A general and versatile synthesis of  $\beta$ -keto esters. *J. Org. Chem.* **43**, 2087-2088 (1978).
45. Davis, J. B., Bailey, J. D. & Sello, J. K. Biomimetic synthesis of a new class of bacterial signaling molecules. *Org. Lett.* **11**, 2984-2987 (2009).
46. Koehn, F. E. & Carter, G. T. The evolving role of natural products in drug discovery. *Nat. Rev. Drug Discov.* **4**, 206-220 (2005).
47. Wolf, E., Kennedy, I. A., Himmeldirk, K. & Spenser, I. D. 5-Hydroxypentane-2,3-dione and 3-amino-1-hydroxypropan-2-one, putative precursors of vitamin B6. *Can. J. Chem.* **75**, 942-948 (1997).
48. Brown, M. D., Gillon, D. W., Meakins, G. D. & Whitham, G. H. A new general synthesis of 2-(N-mono- and N-di-substituted amino)thiazoles. *J. Chem. Soc., Perkin Trans.* **1** 1623-1626 (1985).
49. Edwards, J. A., Guzman, A., Johnson, R., Beeby, P. J. & Fried, J. H. A new total synthesis of ( $\pm$ )-desacetylcephalothin lactone. a synthesis of novel furo[3,4-c]cephams. *Tetrahedron Lett.* **15**, 2031-2034 (1974).
50. Mann, J., Holland, H. J. & Lewis, T. Cycloaddition reactions of polysubstituted furans with oxyallyl carbocations. *Tetrahedron* **43**, 2533-2542 (1987).
51. Pevzner, L. M. Formylation of alkyl cyanofurylmethanphosphonates at the active methylene group. *Russ. J. Gen. Chem.* **83**, 1687-1697 (2013).
52. Morin, J. B., Adams, K. L. & Sello, J. K. Replication of biosynthetic reactions enables efficient synthesis of A-factor, a  $\gamma$ -butyrolactone autoinducer from *Streptomyces griseus*. *Org. Biomol. Chem.* **10**, 1517-1520 (2012).
53. Takano, E. et al. Purification and structural determination of SCB1, a  $\gamma$ -butyrolactone that elicits antibiotic production in *Streptomyces coelicolor* A3(2). *J. Biol. Chem.* **275**, 11010-11016 (2000).
54. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. in *Practical Streptomyces genetics*, The John Innes Foundation, Norwich (2000).
55. Gust, B. et al. Lambda red-mediated genetic manipulation of antibiotic-producing *Streptomyces*. *Adv. Appl. Microbiol.* **54**, 107-128 (2004).



# **Molecular basis for control of antibiotic production**

## **by a bacterial hormone**

Shanshan Zhou,<sup>1†</sup> Hussain Bhukya,<sup>4†</sup> Nicolas Malet,<sup>1†</sup> Peter J. Harrison,<sup>1</sup> Dean Rea,<sup>2</sup>  
Matthew J. Belousoff,<sup>5</sup> Hariprasad Venugopal,<sup>6</sup> Paulina K. Sydor,<sup>1</sup> Kathryn M. Styles,<sup>2</sup>  
Lijiang Song,<sup>1</sup> Max J. Cryle,<sup>4,7</sup> Lona M. Alkhalaf,<sup>1</sup> Vilmos Fülöp,<sup>2</sup> Gregory L. Challis<sup>1,3,4,7\*</sup> and  
Christophe Corre<sup>1,2,3\*</sup>

<sup>1</sup>Department of Chemistry, University of Warwick, Coventry CV4 7AL, UK

<sup>2</sup>School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK

<sup>3</sup>Warwick Integrative Synthetic Biology Centre, University of Warwick, Coventry CV4 7AL, UK

<sup>4</sup>Department of Biochemistry and Molecular Biology, Monash University, Clayton VIC 3800, Australia.

<sup>5</sup>Infection and Immunity Program, Department of Microbiology, Monash University, Clayton, VIC 3800, Australia.

<sup>6</sup>Ramaciotti Centre for Electron Microscopy, Monash University, Clayton, VIC 3800, Australia.

<sup>7</sup>ARC Centre of Excellence for Innovations in Peptide and Protein Chemistry, Monash University, Clayton VIC 3800, Australia.

<sup>†</sup> These authors contributed equally

\*Corresponding authors:

Dr C. Corre, School of Life Sciences, University of Warwick, Coventry CV4 7AL, U.K. Tel 44-2476-523557. Email [C.Corre@warwick.ac.uk](mailto:C.Corre@warwick.ac.uk)

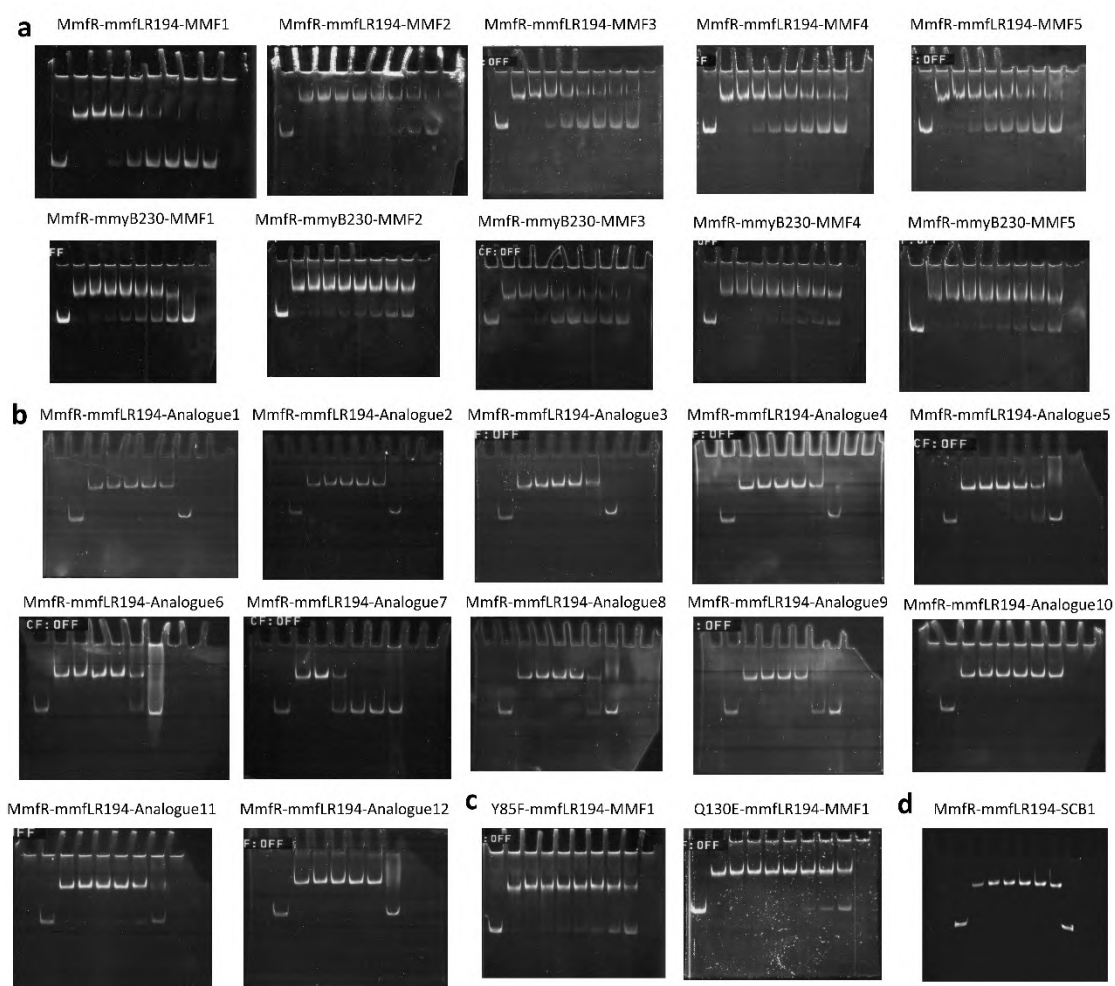
Prof G.L. Challis, Department of Chemistry, University of Warwick, Coventry CV4 7AL, U.K. Tel 44-2476-574024. Email [G.L.Challis@warwick.ac.uk](mailto:G.L.Challis@warwick.ac.uk)

**-Supplementary Information-**

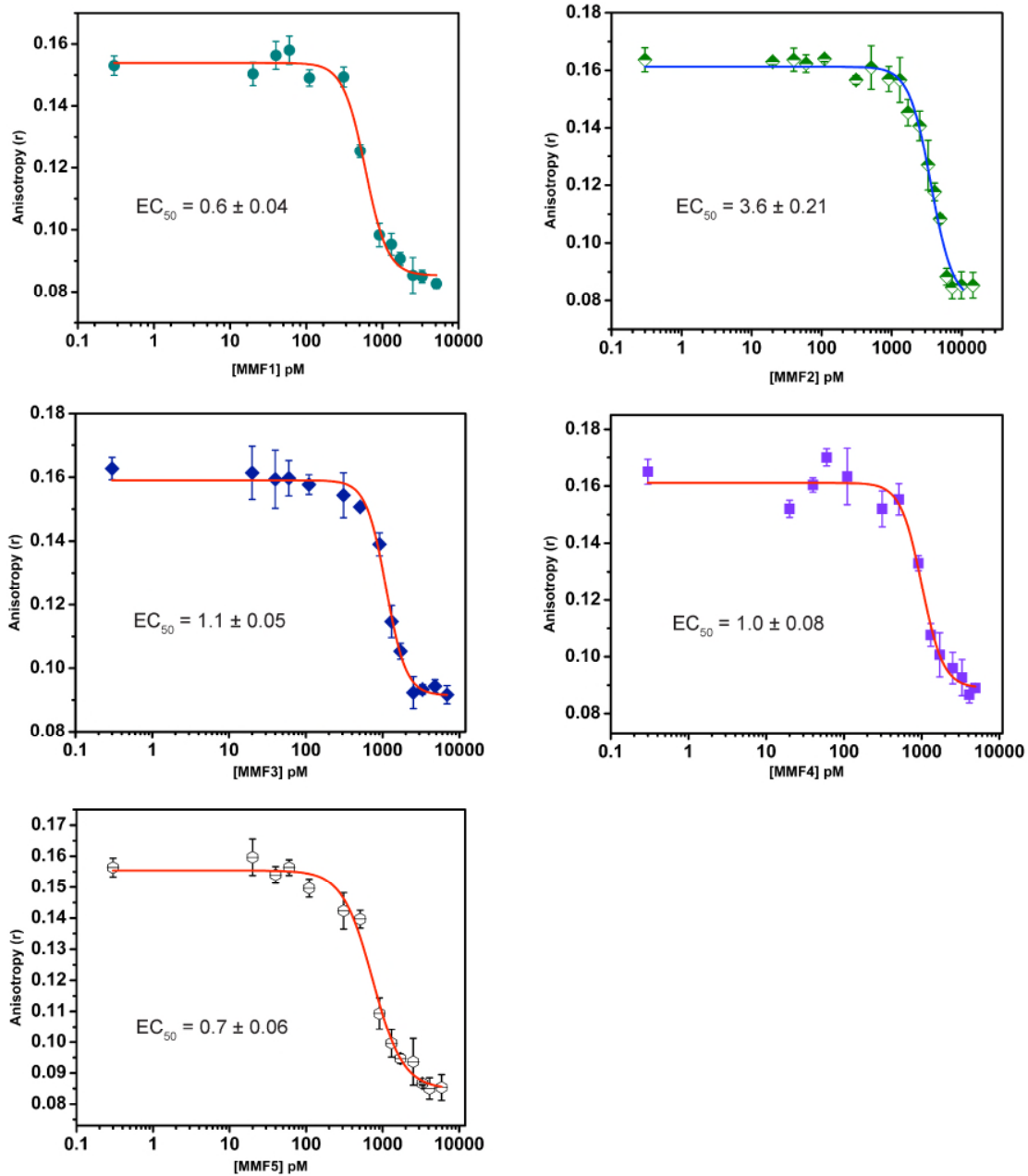
## Table of contents

### Supplementary Figures

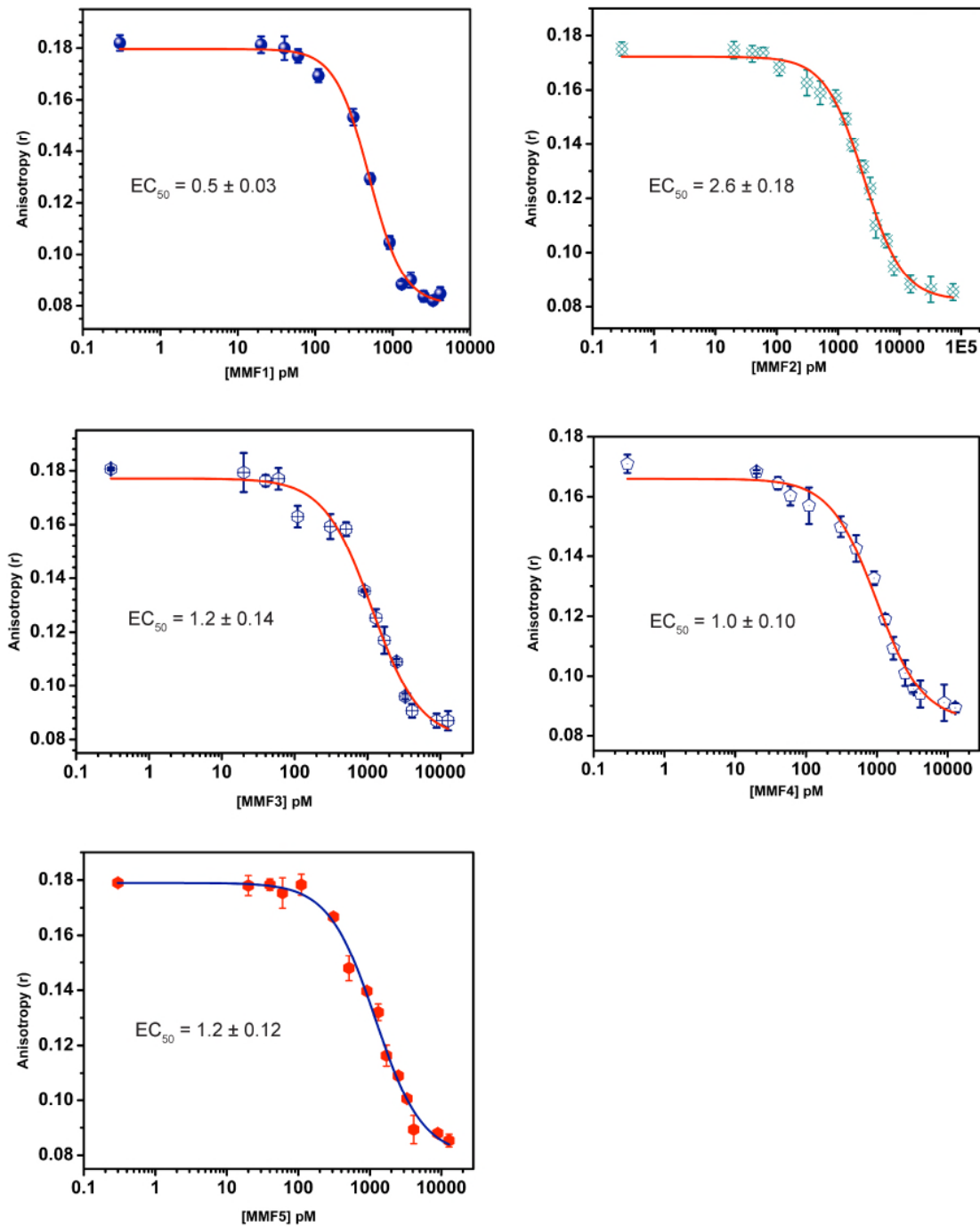
<b>Supplementary Fig. 1</b>   Raw data and gel images	S3
<b>Supplementary Fig. 2</b>   FA plots for release of MmfR from MARE1 by MMFs 1-5	S4
<b>Supplementary Fig. 3</b>   FA plots for release of MmfR from MARE2 by MMFs 1-5	S5
<b>Supplementary Fig. 4</b>   FA plots for binding of MmfR mutants to MARE1 and subsequent release by MMF1	S6
<b>Supplementary Fig. 5-7</b>   <sup>1</sup> H-NMR, <sup>13</sup> C-NMR and MS spectra of analogue 2	S7
<b>Supplementary Fig. 8-10</b>   <sup>1</sup> H-NMR, <sup>13</sup> C-NMR and MS spectra of analogue 3	S9
<b>Supplementary Fig. 11-13</b>   <sup>1</sup> H-NMR, <sup>13</sup> C-NMR and MS spectra of analogue 4	S11
<b>Supplementary Fig. 14-16</b>   <sup>1</sup> H-NMR, <sup>13</sup> C-NMR and MS spectra of analogue 5	S13
<b>Supplementary Fig. 17-19</b>   <sup>1</sup> H-NMR, <sup>13</sup> C-NMR and MS spectra of analogue 6	S15
<b>Supplementary Fig. 20-22</b>   <sup>1</sup> H-NMR, <sup>13</sup> C-NMR and MS spectra of analogue 7	S17
<b>Supplementary Fig. 23-25</b>   <sup>1</sup> H-NMR, <sup>13</sup> C-NMR and MS spectra of analogue 8	S19
<b>Supplementary Fig. 26-28</b>   <sup>1</sup> H-NMR, <sup>13</sup> C-NMR and MS spectra of analogue 9	S21
<b>Supplementary Fig. 29-31</b>   <sup>1</sup> H-NMR, <sup>13</sup> C-NMR and MS spectra of analogue 11	S23
<b>Supplementary Fig. 32-34</b>   <sup>1</sup> H-NMR, <sup>13</sup> C-NMR and MS spectra of analogue 12	S25
<b>Supplementary Fig. 35</b>   FA plots for release of MmfR from MARE1 by selected analogues	S27



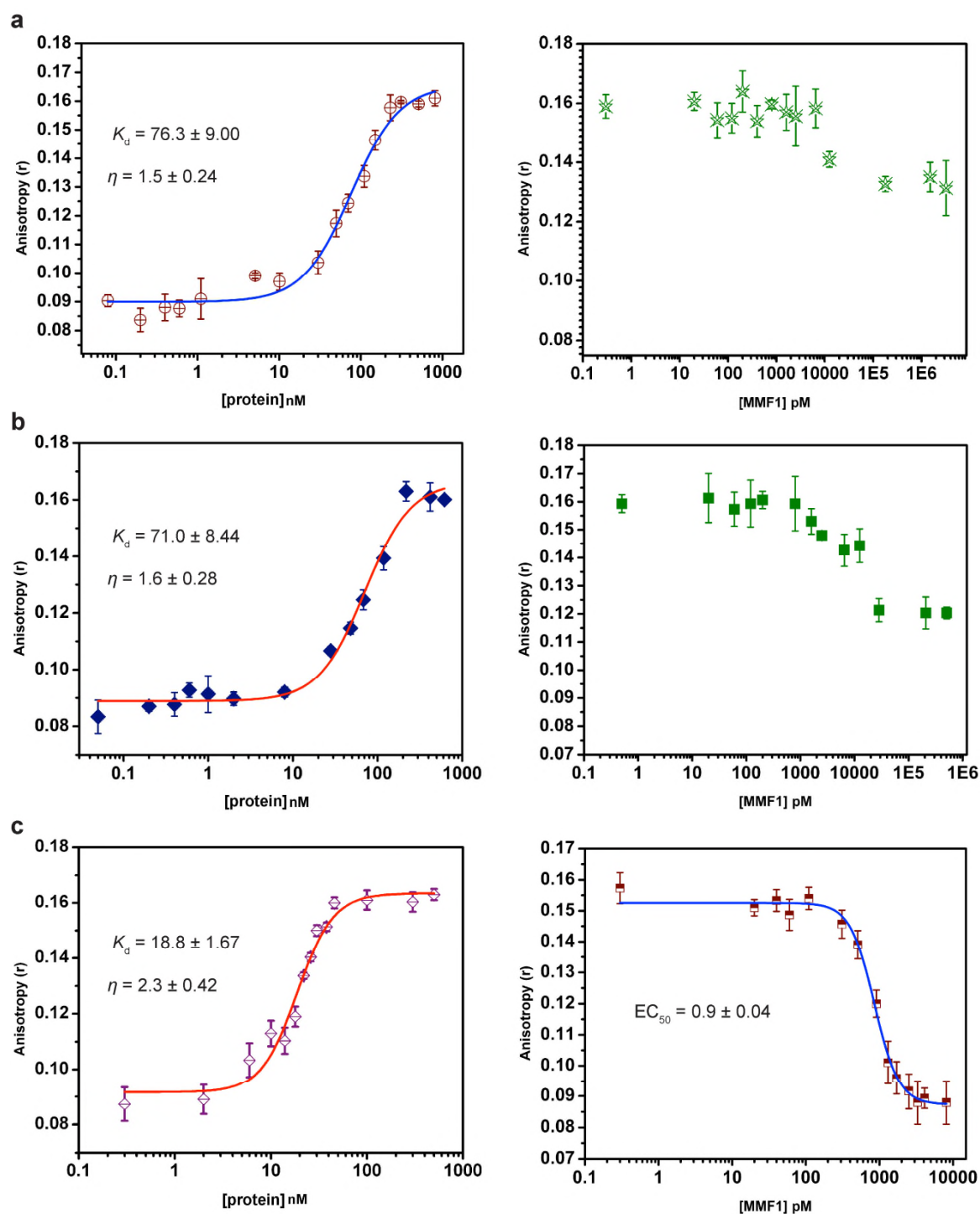
**Supplementary Fig. 1. | Raw data and gel images.** **a**, Original EMSA gels for Extended Data Figures 3a and 3b. **b**, Original EMSA gels for Extended Data Figures 3e and 10. **c**, Original EMSA gels for Extended Data Figure 3d. **d**, Original EMSA gel for Extended Data Figure 3c. In these EMSAs, lane 1 from the left only contains the corresponding isolated DNA fragments, which is used as a reference for the released DNA by ligands from the DNA-protein complex. The gel images were cropped for the final images and the color inverted to show the DNA bands in black in white background. The right most lane in the gel images in **b** contains 1000 nmol of compounds, which cause protein precipitation and thus be cropped from the final image. This is the same case for EMSA gel in **d**, in which the right most lane contained 4000 nmol of SCB1 and omitted from the final image. At least two independent technical replicates of each EMSA were conducted and in all cases similar results were obtained.



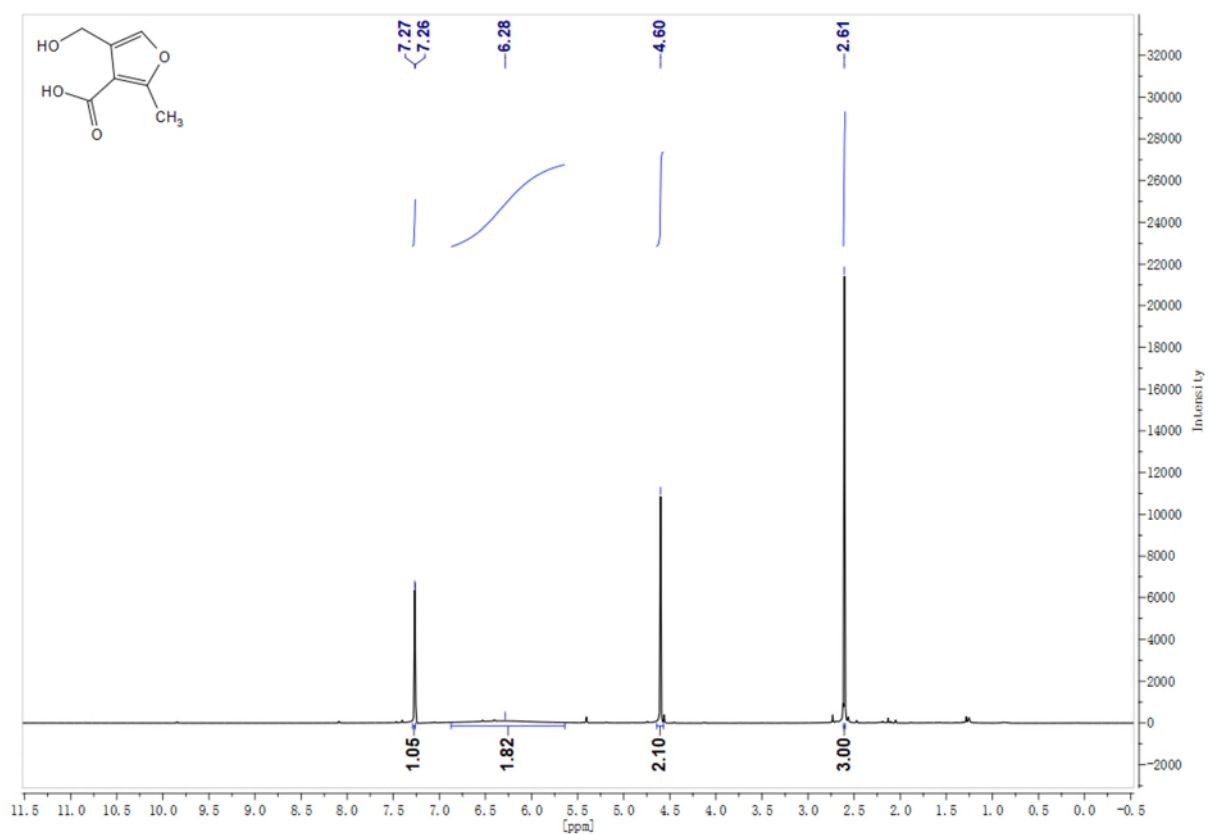
**Supplementary Fig. 2 | FA plots for release of MmFR from MARE1 by MMFs 1-5.** Data points are the mean of three independent technical replicates (n=3) and error bars represent  $\pm 1$  standard deviation. The  $EC_{50}$  (nM) calculated from each data set is shown.



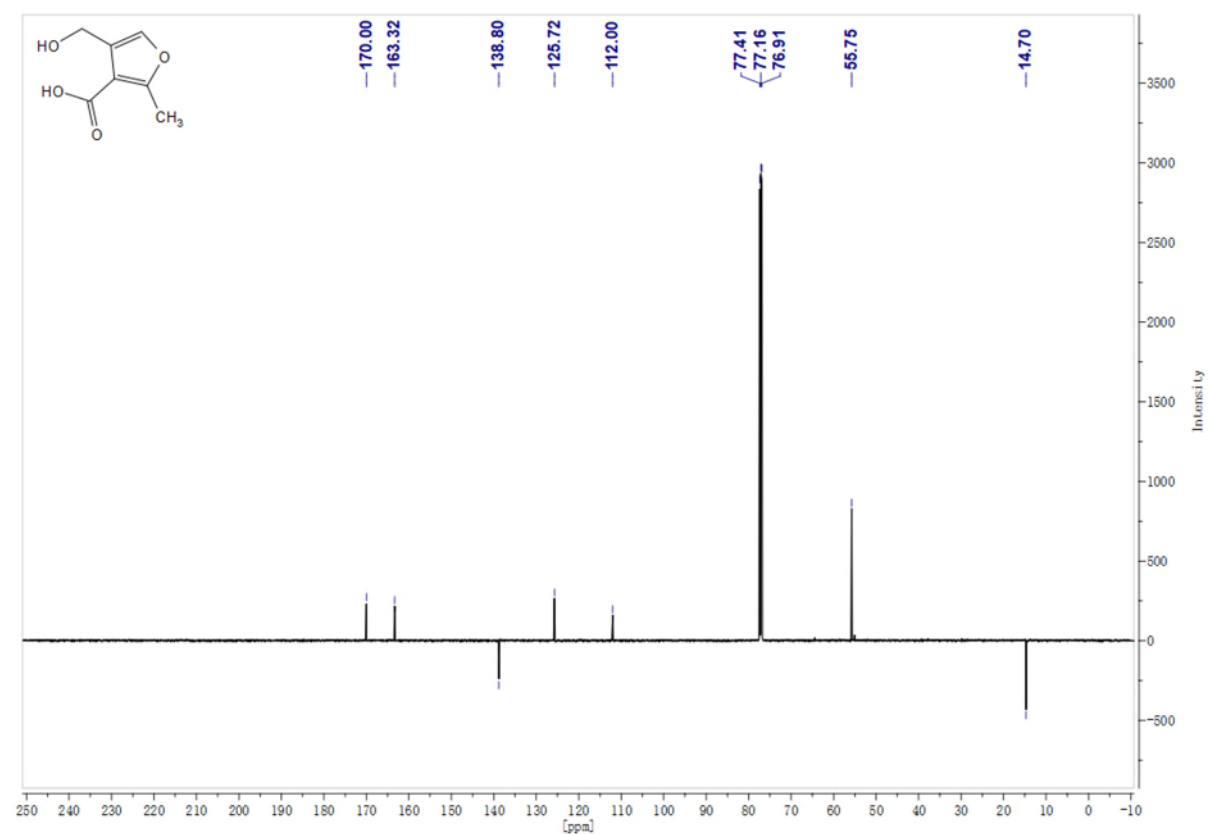
**Supplementary Fig. 3 | FA plots for release of Mmfr from MARE2 by MMFs 1-5.** Data points are the mean of three independent technical replicates (n=3) and error bars represent  $\pm 1$  standard deviation. The  $EC_{50}$  (nM) calculated from each data set is shown.



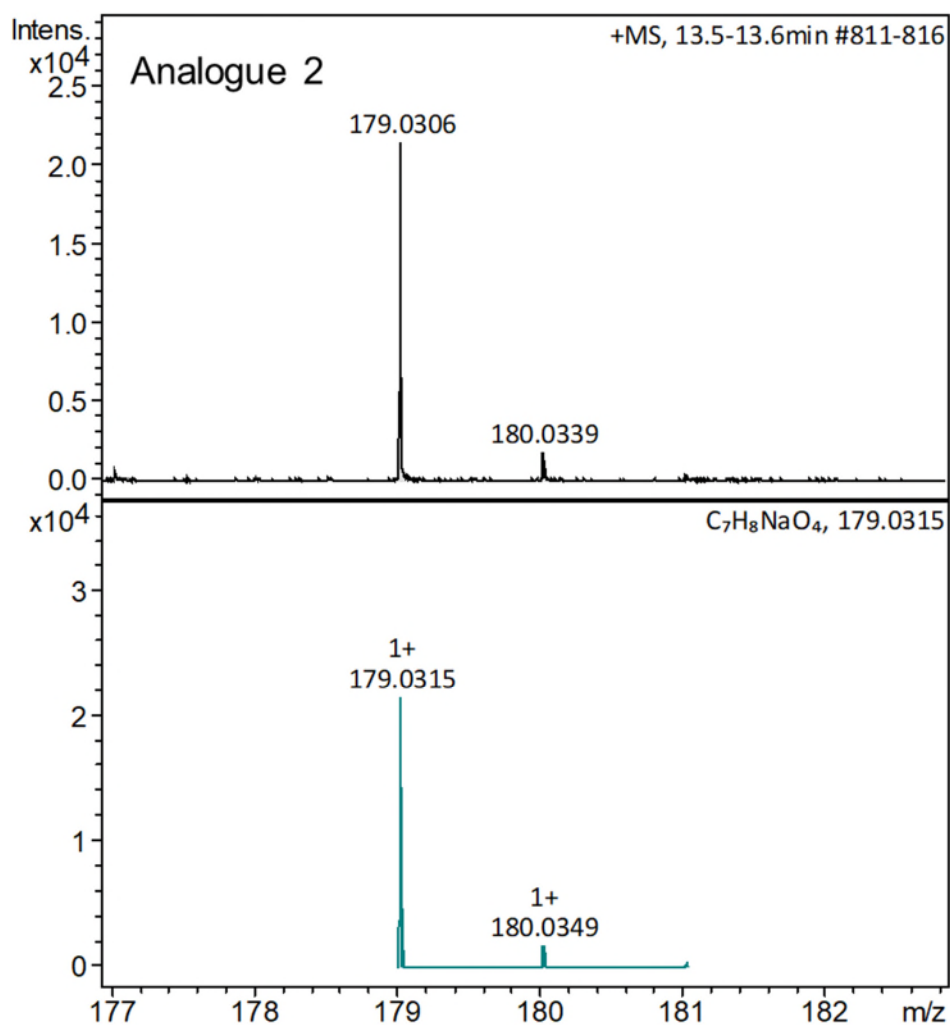
**Supplementary Fig. 4 | FA plots for binding of MmfR mutants to MARE1 and subsequent release by MMF1.** Data points are the mean of three independent technical replicates (n=3) and error bars represent  $\pm 1$  standard deviation. Binding is shown in the left panel and release in the right panel. The  $K_d$  (nM) and Hill coefficient ( $\eta$ ) calculated from each binding data set and the  $EC_{50}$  (nM) calculated from each release data set are shown. **a**, Y85F mutant. **b**, Q130E mutant. **c**, Y144F mutant.



Supplementary Fig. 5 | <sup>1</sup>H-NMR spectrum of analogue 2.

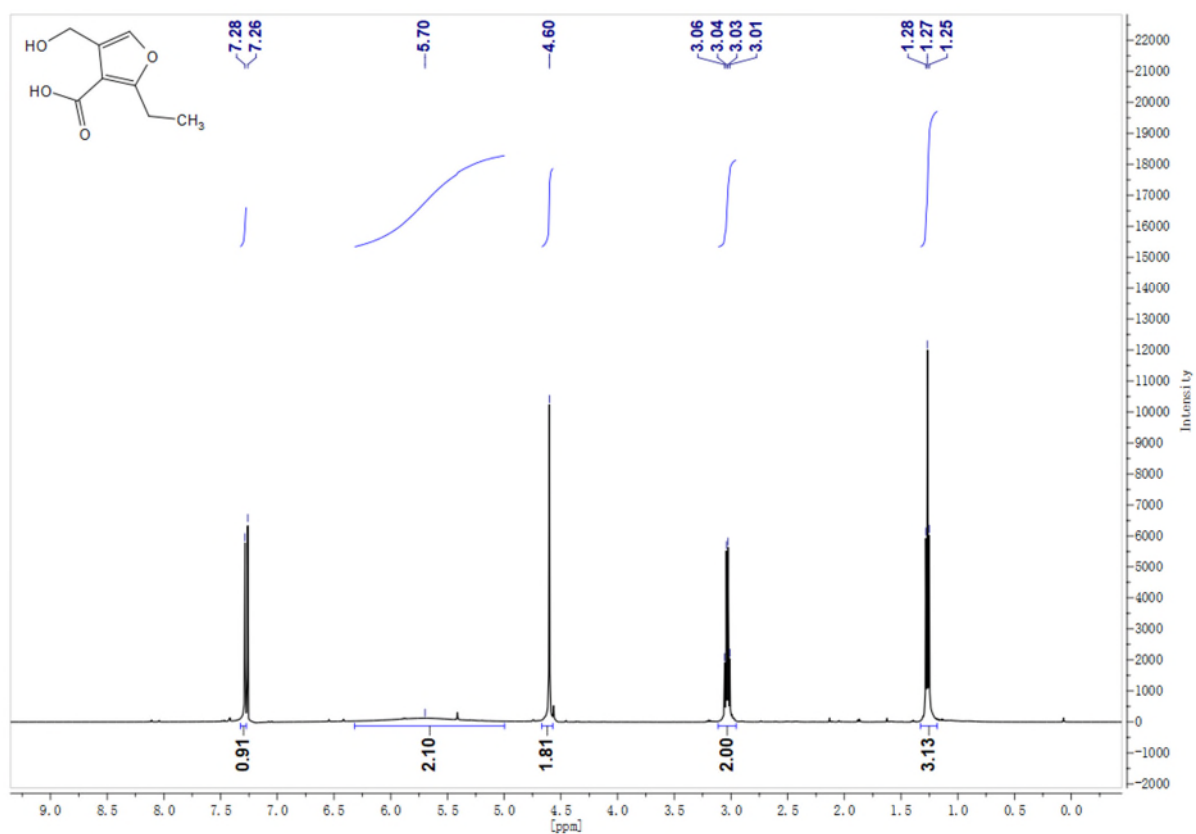


Supplementary Fig. 6 | <sup>13</sup>C-NMR spectrum of analogue 2.

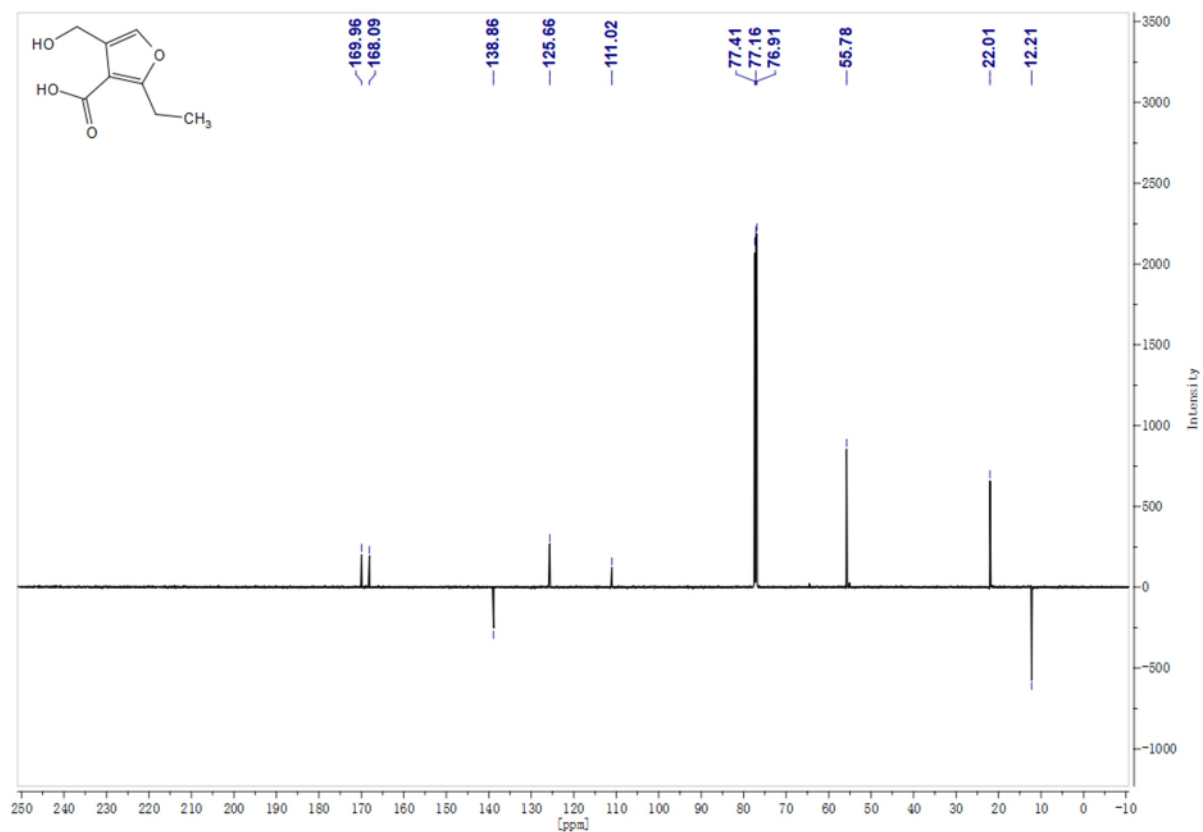


**Supplementary Fig. 7 | Comparison of measured (top panel) and simulated (bottom panel) mass spectra for analogue 2.**

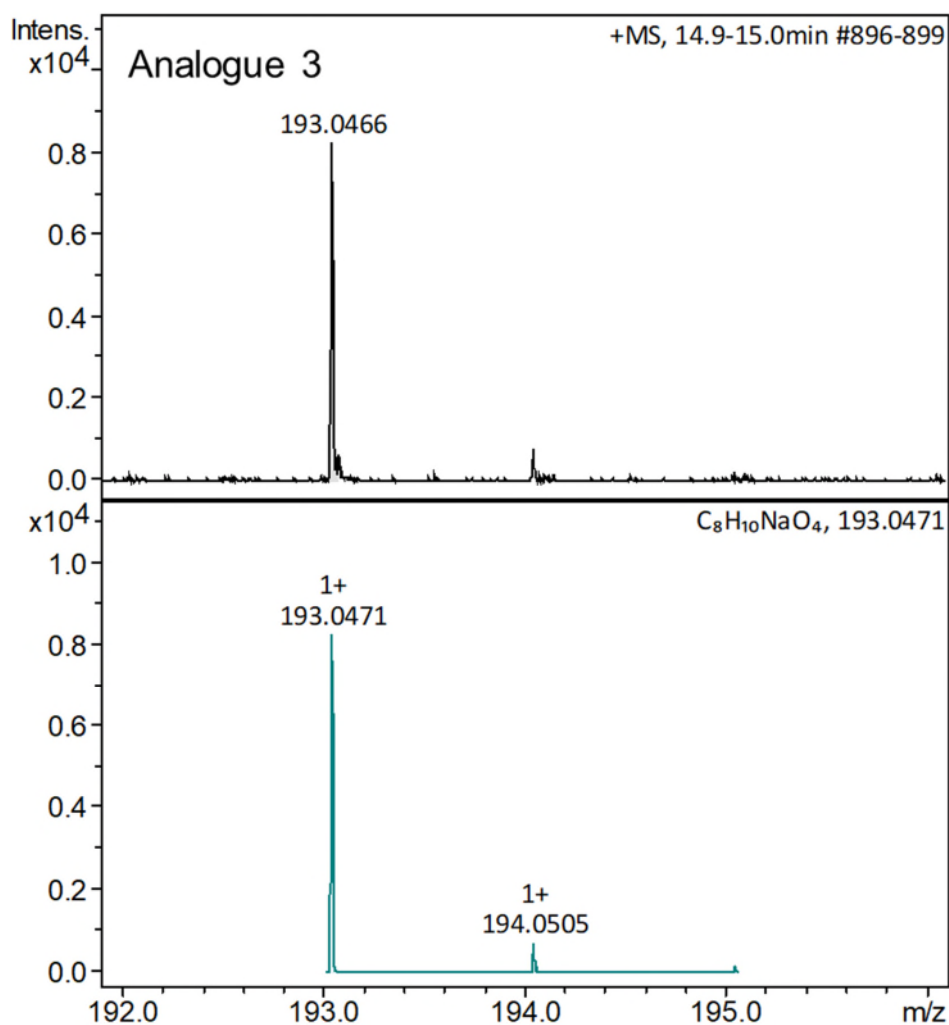




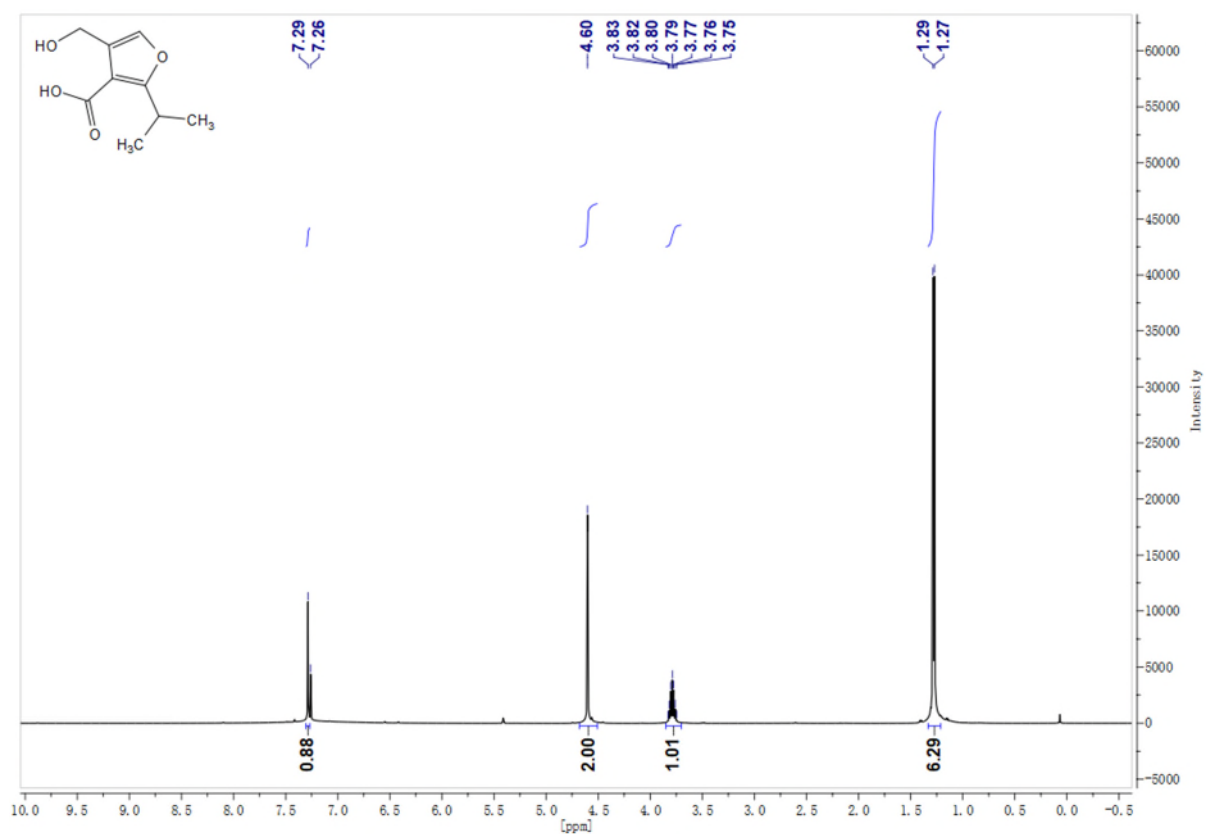
Supplementary Fig. 8 | <sup>1</sup>H-NMR spectrum of analogue 3.



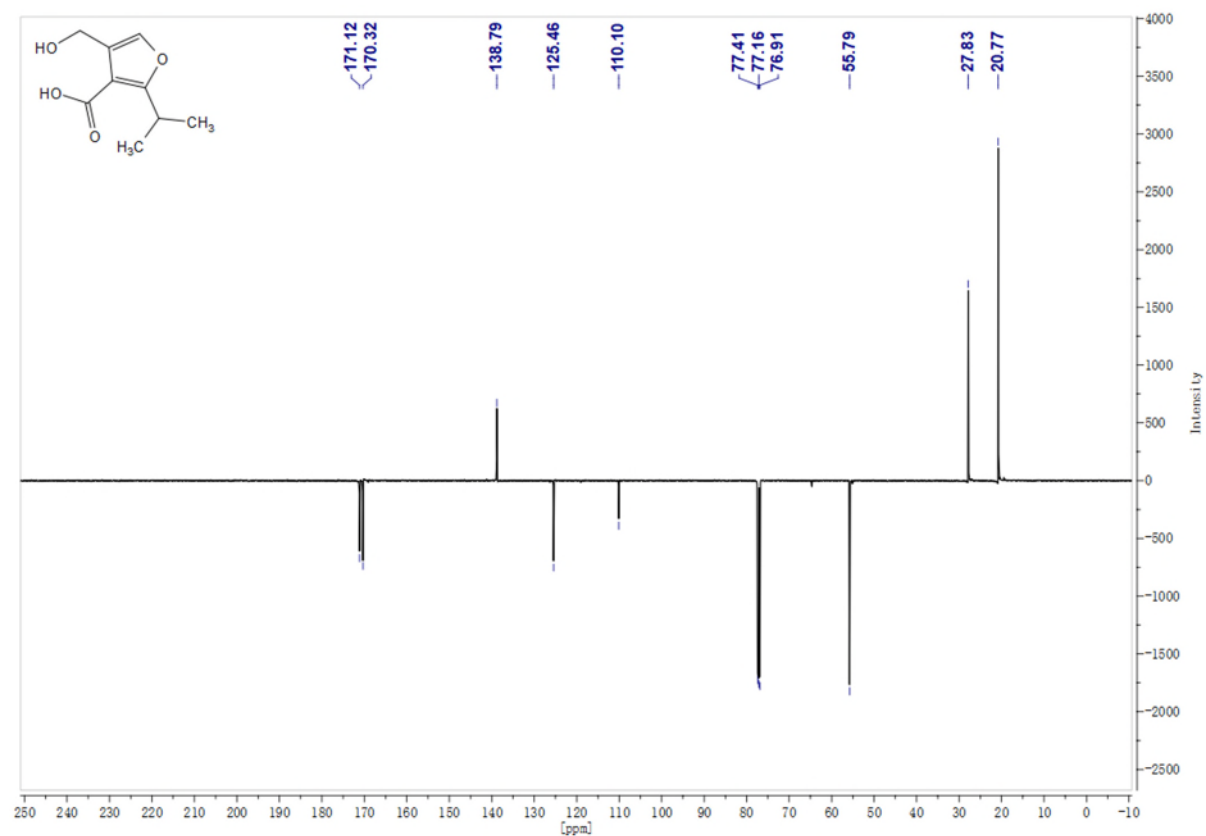
Supplementary Fig. 9 | <sup>13</sup>C-NMR spectrum of analogue 3.



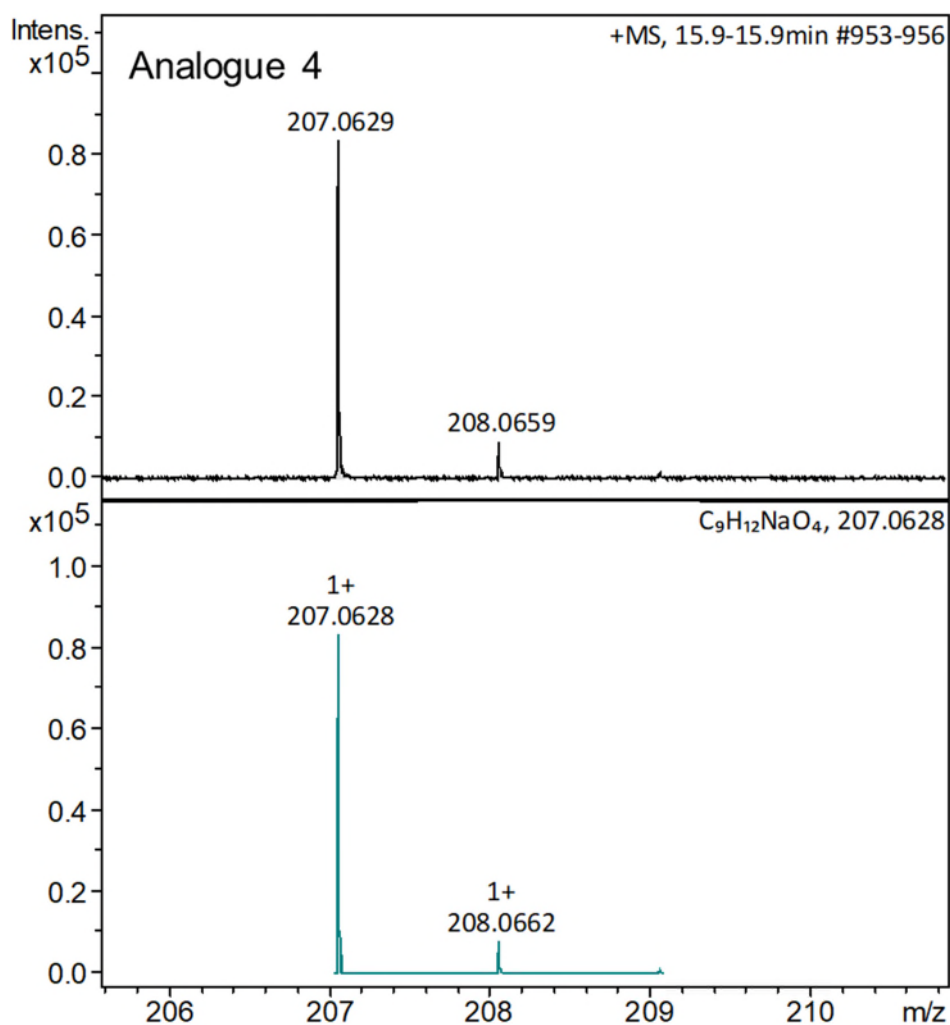
**Supplementary Fig. 10 | Comparison of measured (top panel) and simulated (bottom panel) mass spectra for analogue 3.**



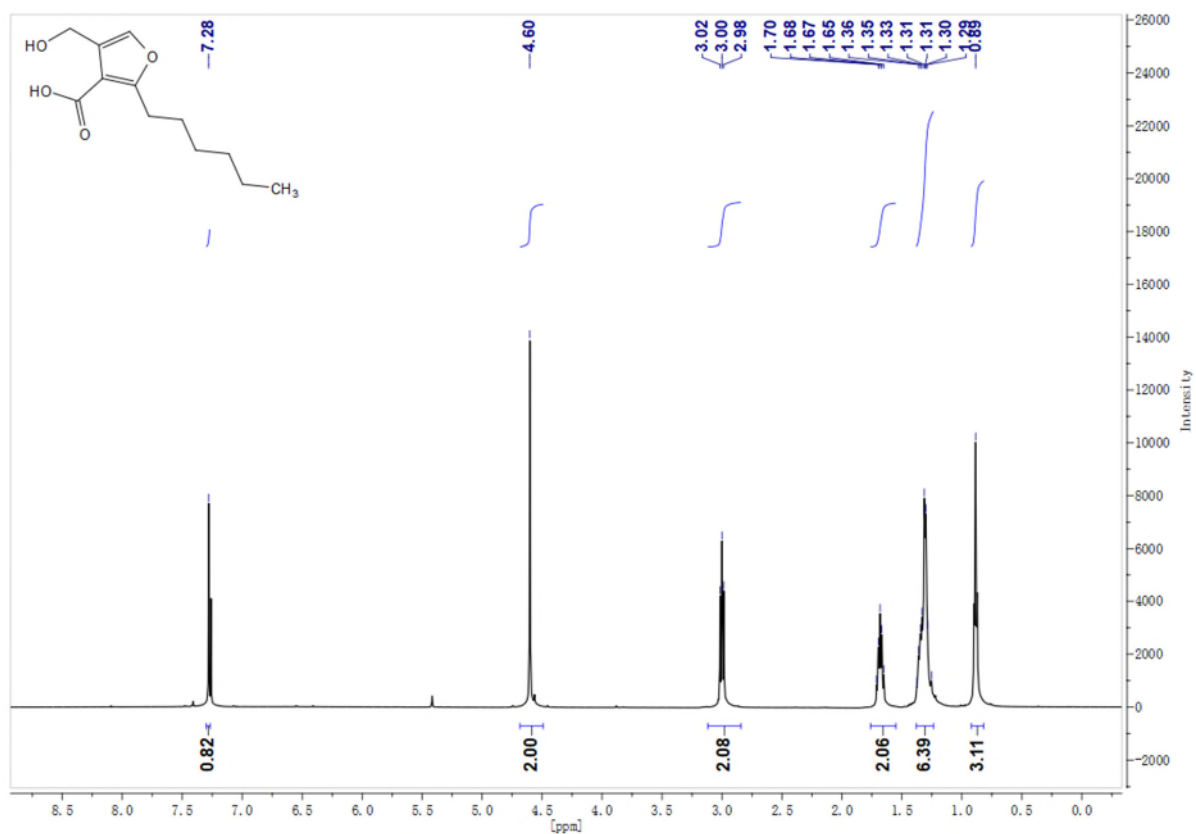
Supplementary Fig. 11 | <sup>1</sup>H-NMR spectrum of analogue 4.



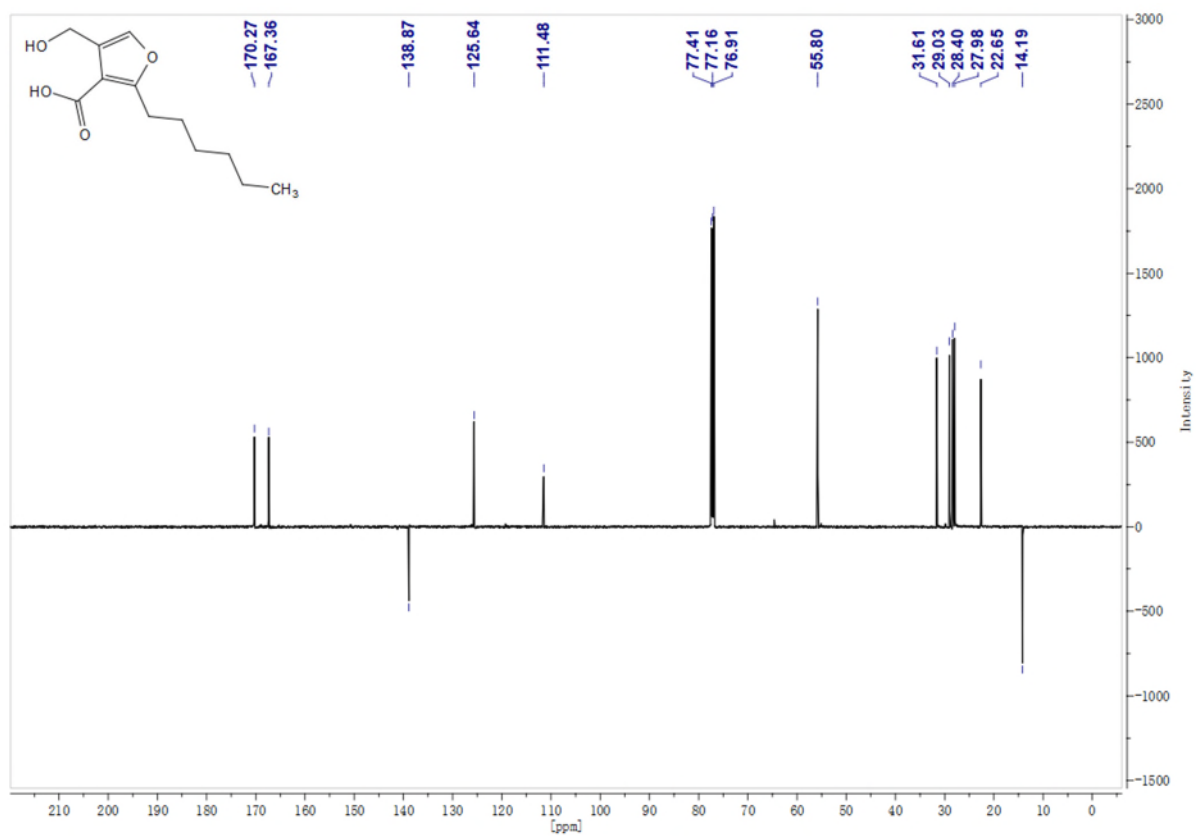
Supplementary Fig. 12 | <sup>13</sup>C-NMR spectrum of analogue 4.



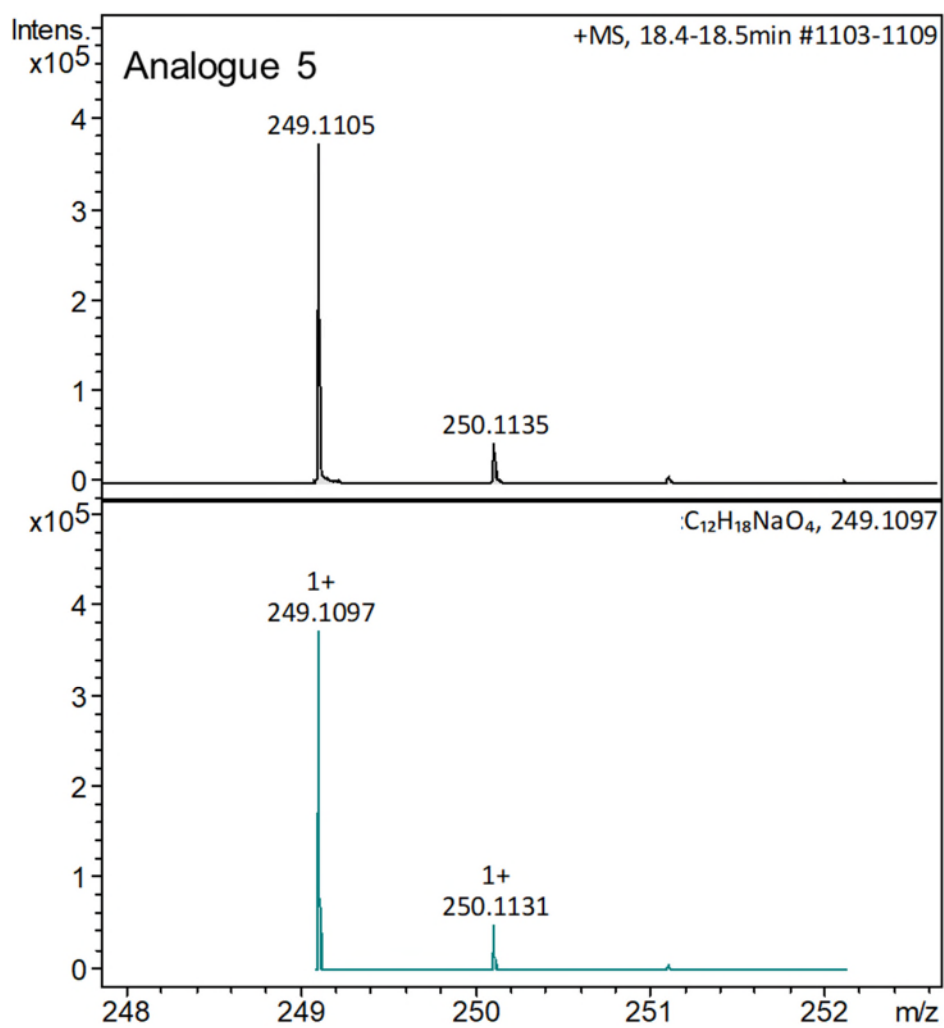
**Supplementary Fig. 13 | Comparison of measured (top panel) and simulated (bottom panel) mass spectra for analogue 4.**



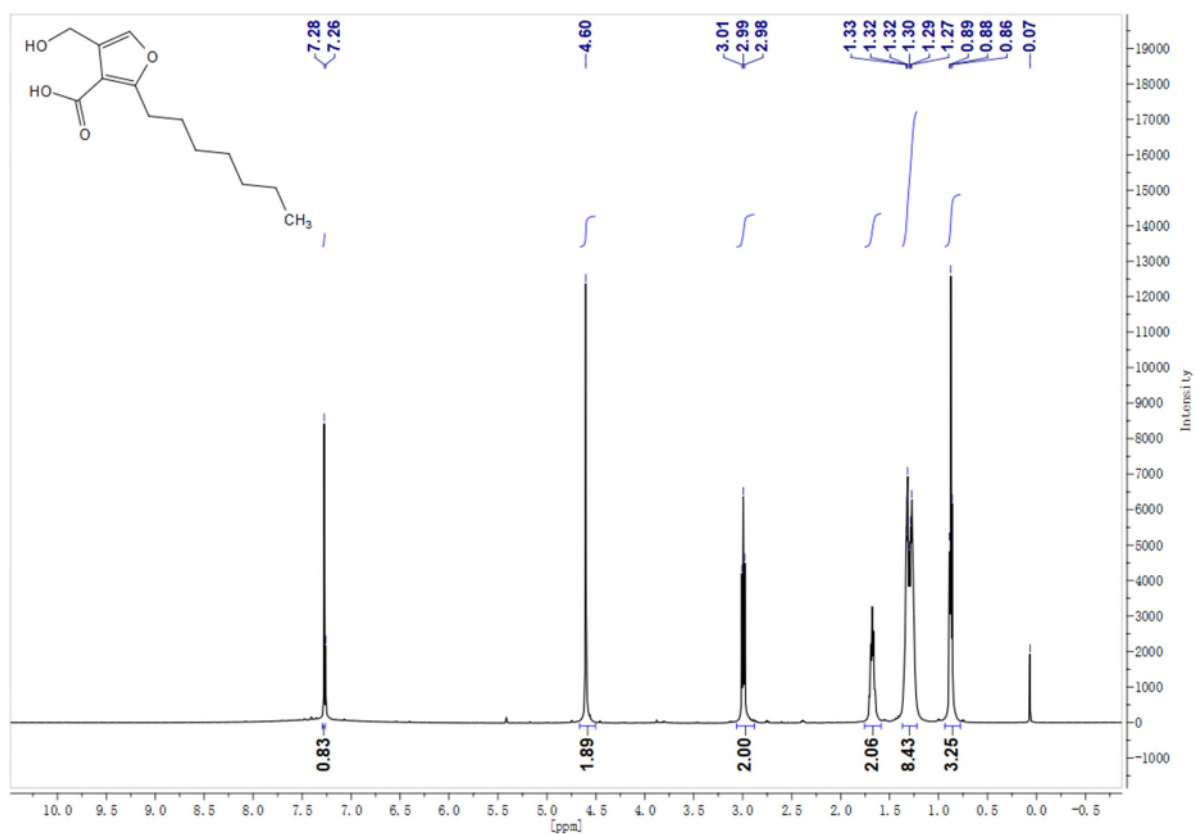
Supplementary Fig. 14 | <sup>1</sup>H-NMR spectrum of analogue 5.



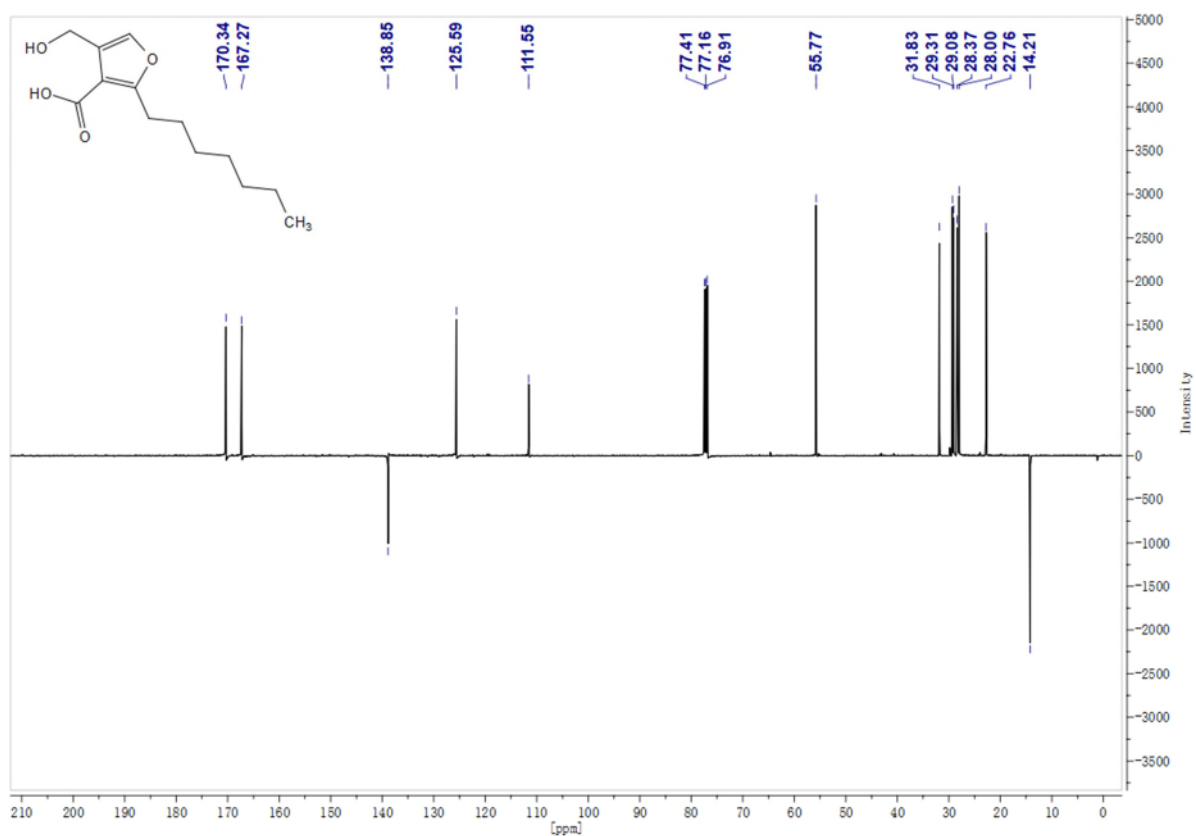
Supplementary Fig. 15 | <sup>13</sup>C-NMR spectrum of analogue 5.



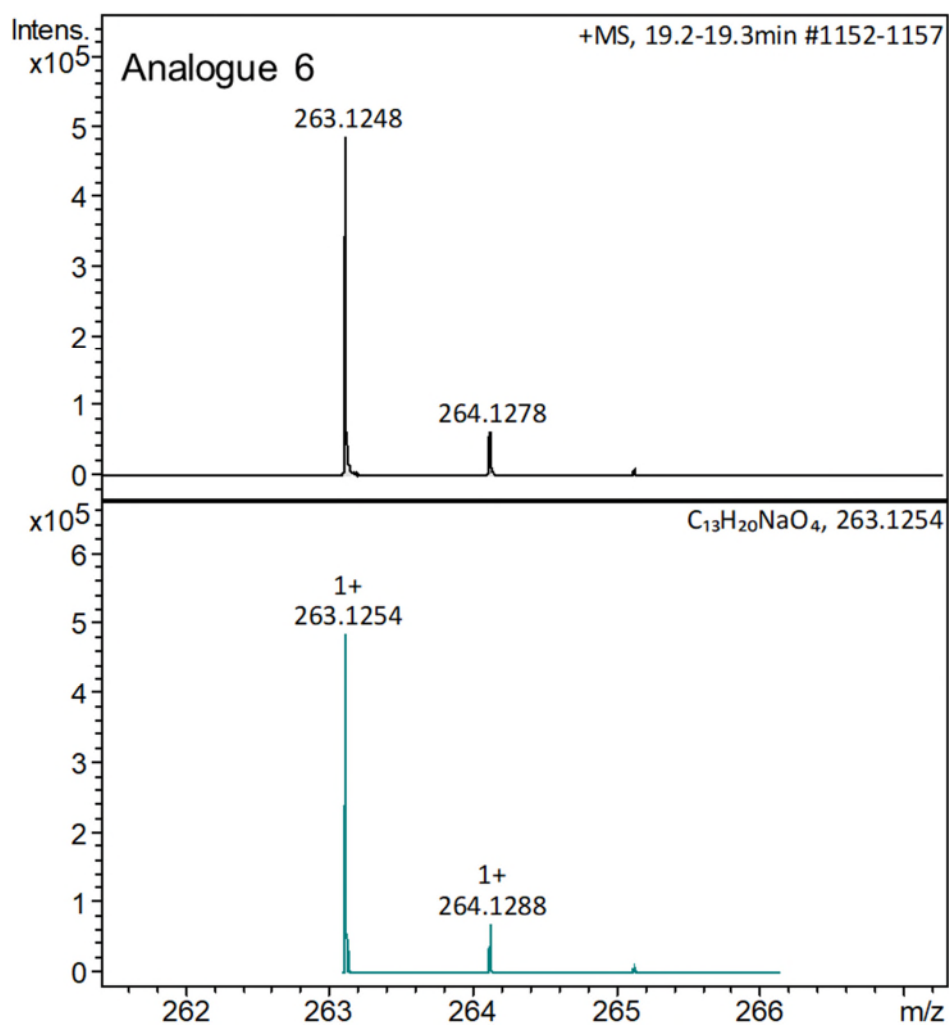
**Supplementary Fig. 16 | Comparison of measured (top panel) and simulated (bottom panel) mass spectra for analogue 5.**



Supplementary Fig. 17 | <sup>1</sup>H-NMR spectrum of analogue 6.

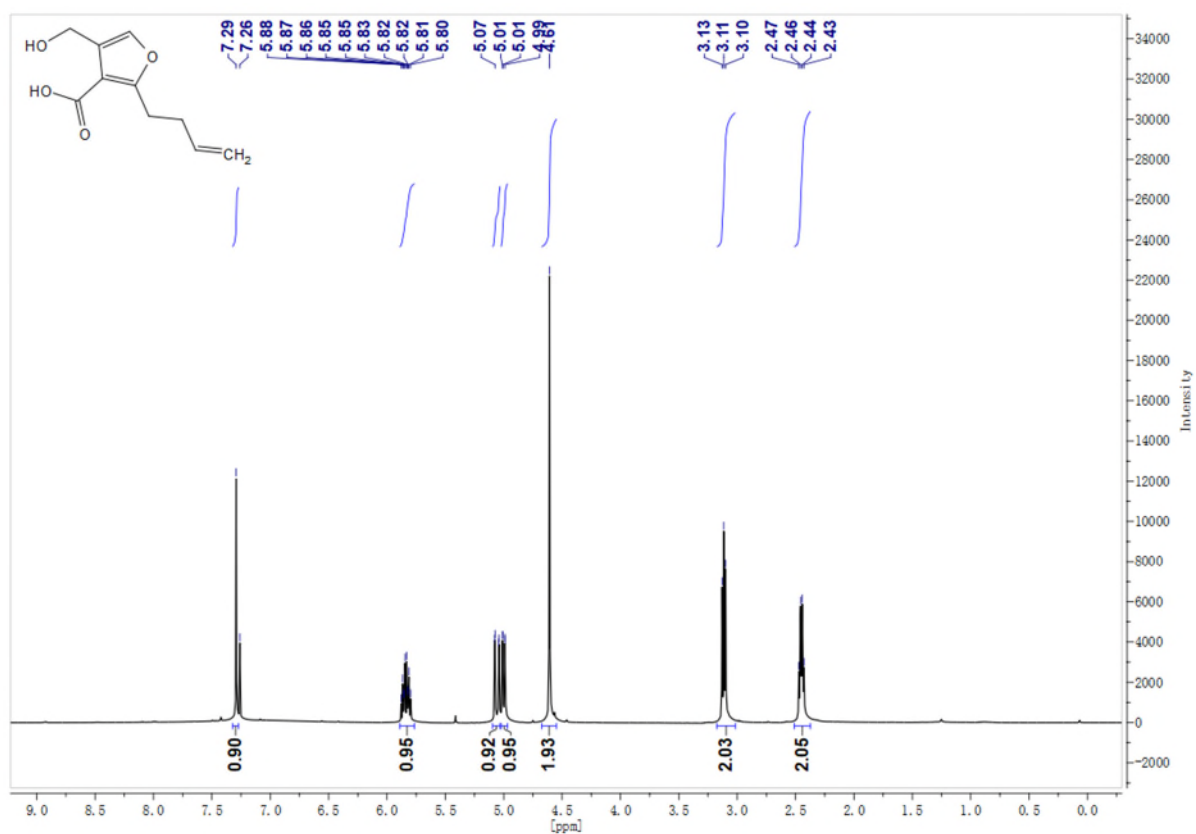


Supplementary Fig. 18 | <sup>13</sup>C-NMR spectrum of analogue 6.

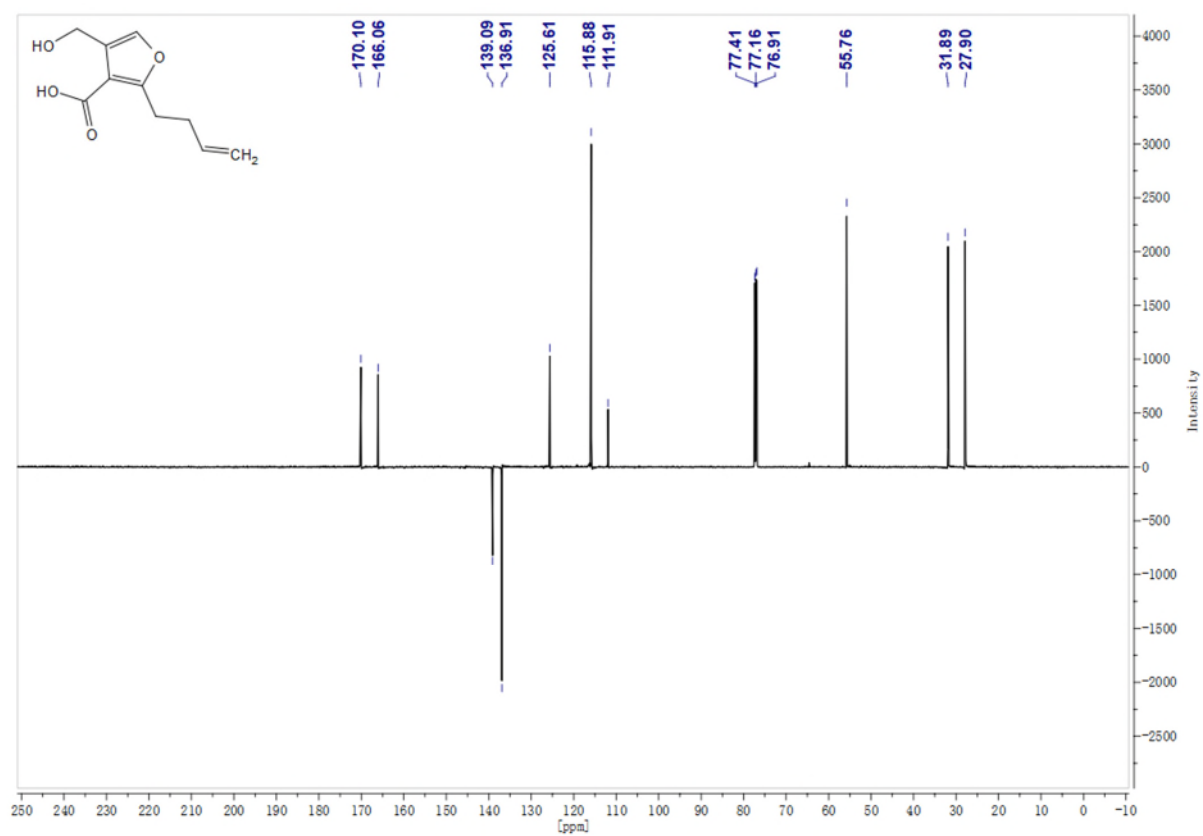


**Supplementary Fig. 19 | Comparison of measured (top panel) and simulated (bottom panel) mass spectra for analogue 6.**

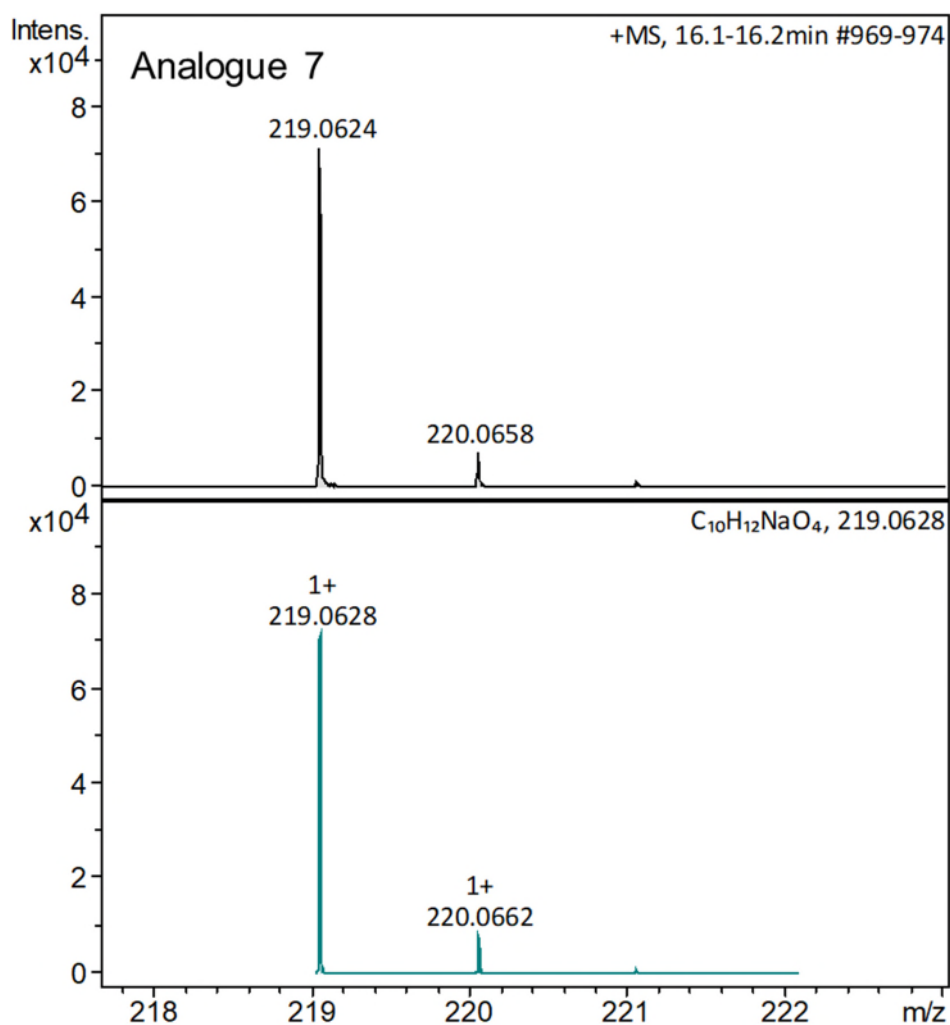




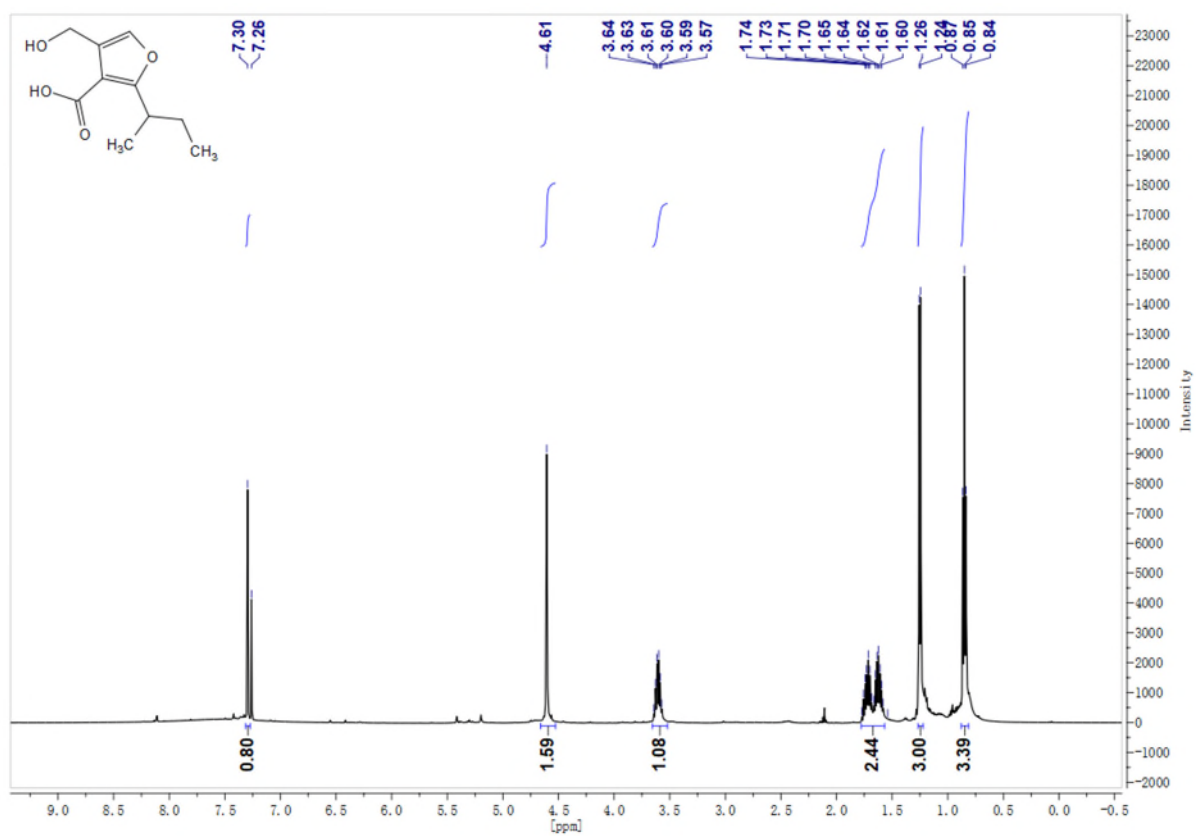
Supplementary Fig. 20 | <sup>1</sup>H-NMR spectrum of analogue 7.



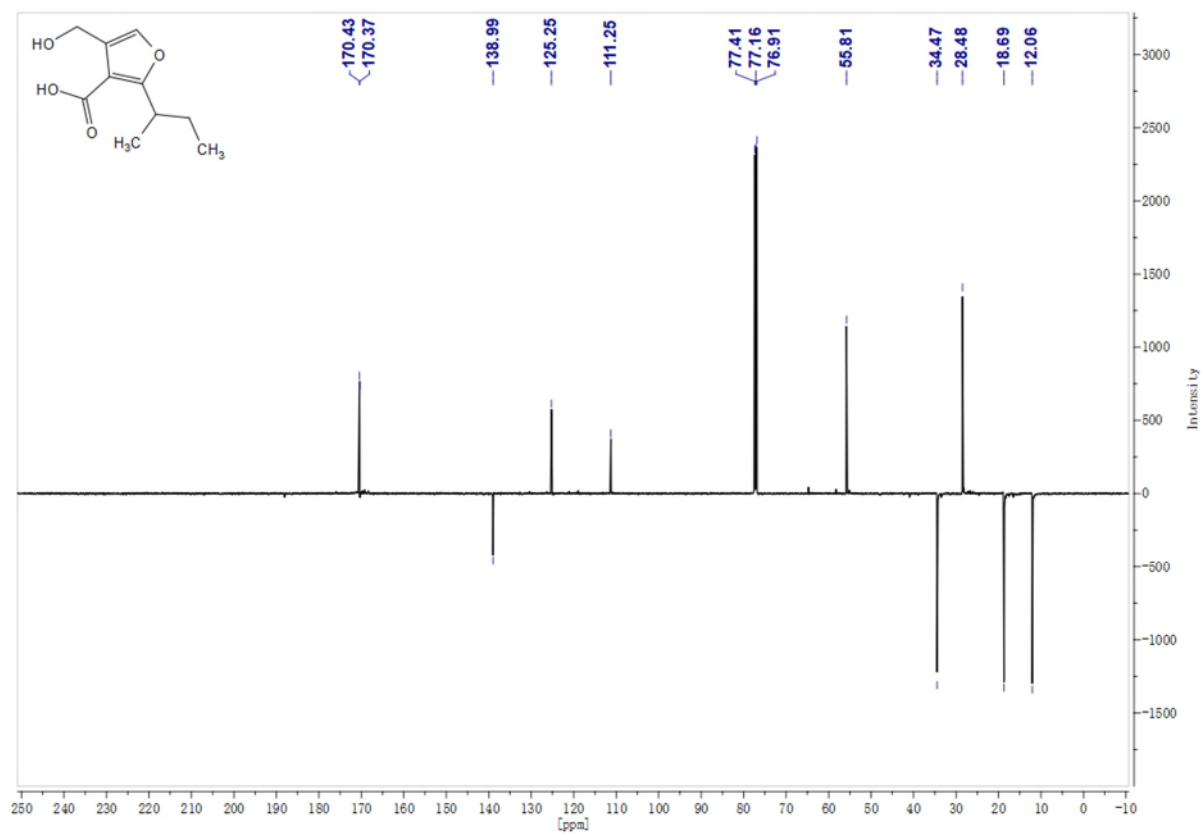
Supplementary Fig. 21 | <sup>13</sup>C-NMR spectrum of analogue 7.



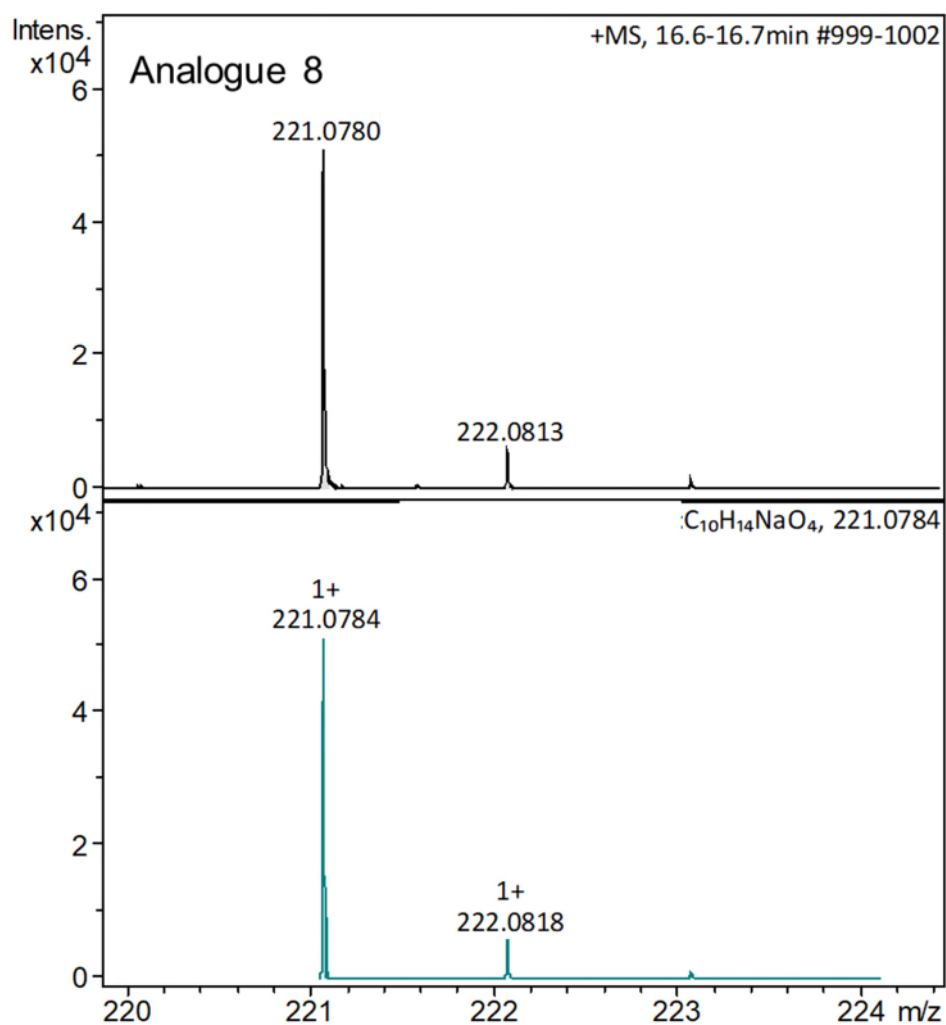
**Supplementary Fig. 22 | Comparison of measured (top panel) and simulated (bottom panel) mass spectra for analogue 7.**



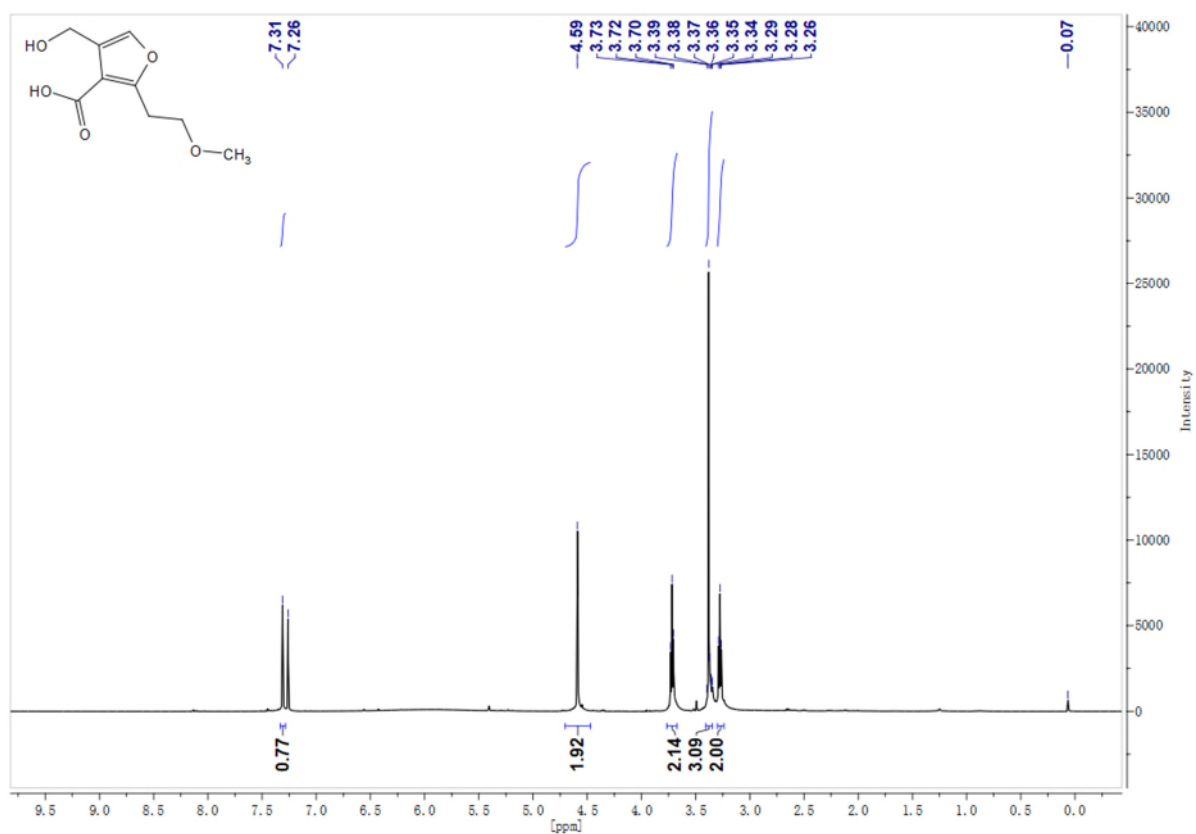
Supplementary Fig. 23 | <sup>1</sup>H-NMR spectrum of analogue 8.



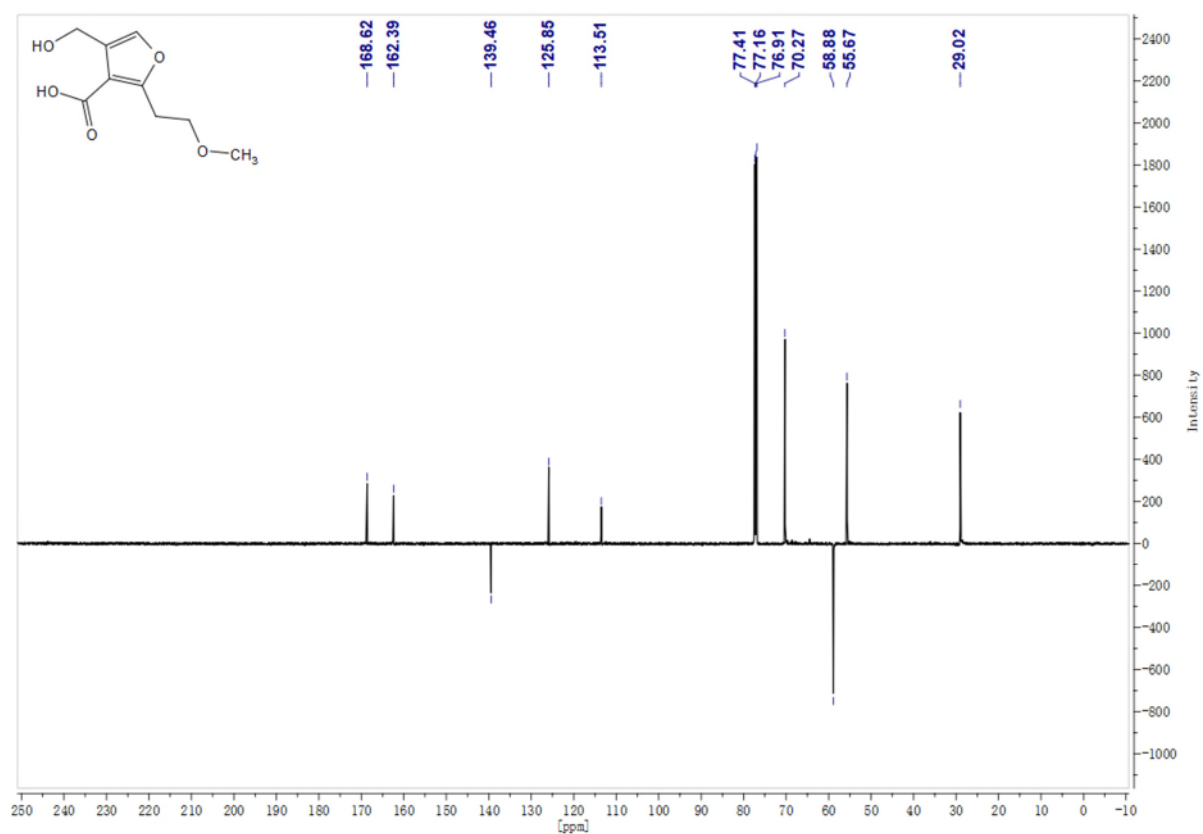
Supplementary Fig. 24 | <sup>13</sup>C-NMR spectrum of analogue 8.



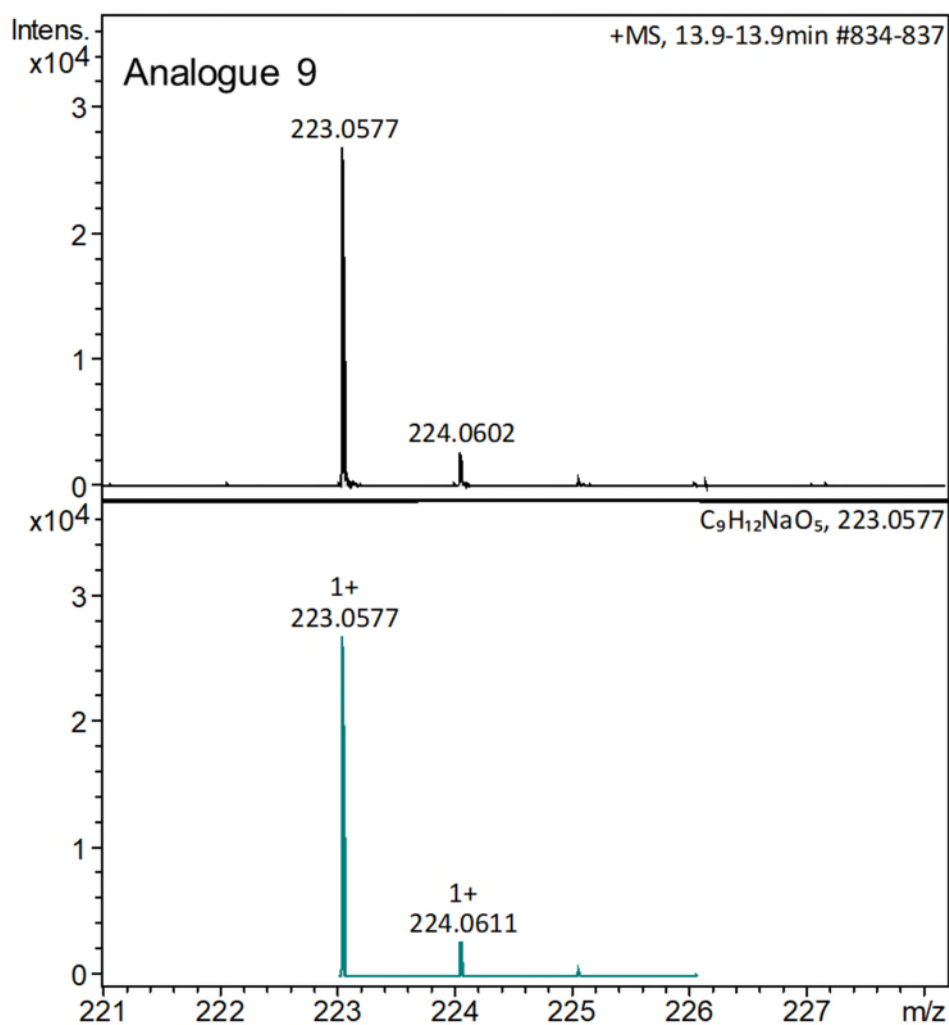
**Supplementary Fig. 25 | Comparison of measured (top panel) and simulated (bottom panel) mass spectra for analogue 8.**



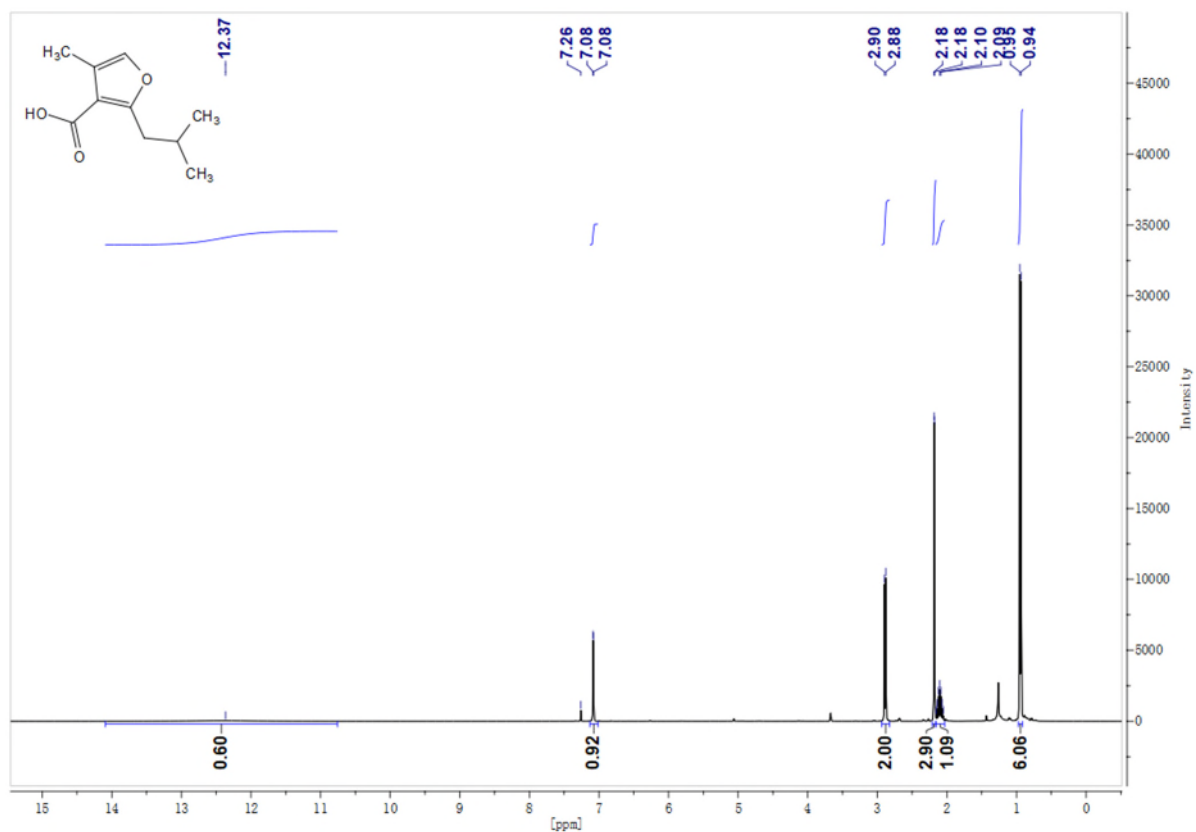
Supplementary Fig. 26 | <sup>1</sup>H-NMR spectrum of analogue 9.



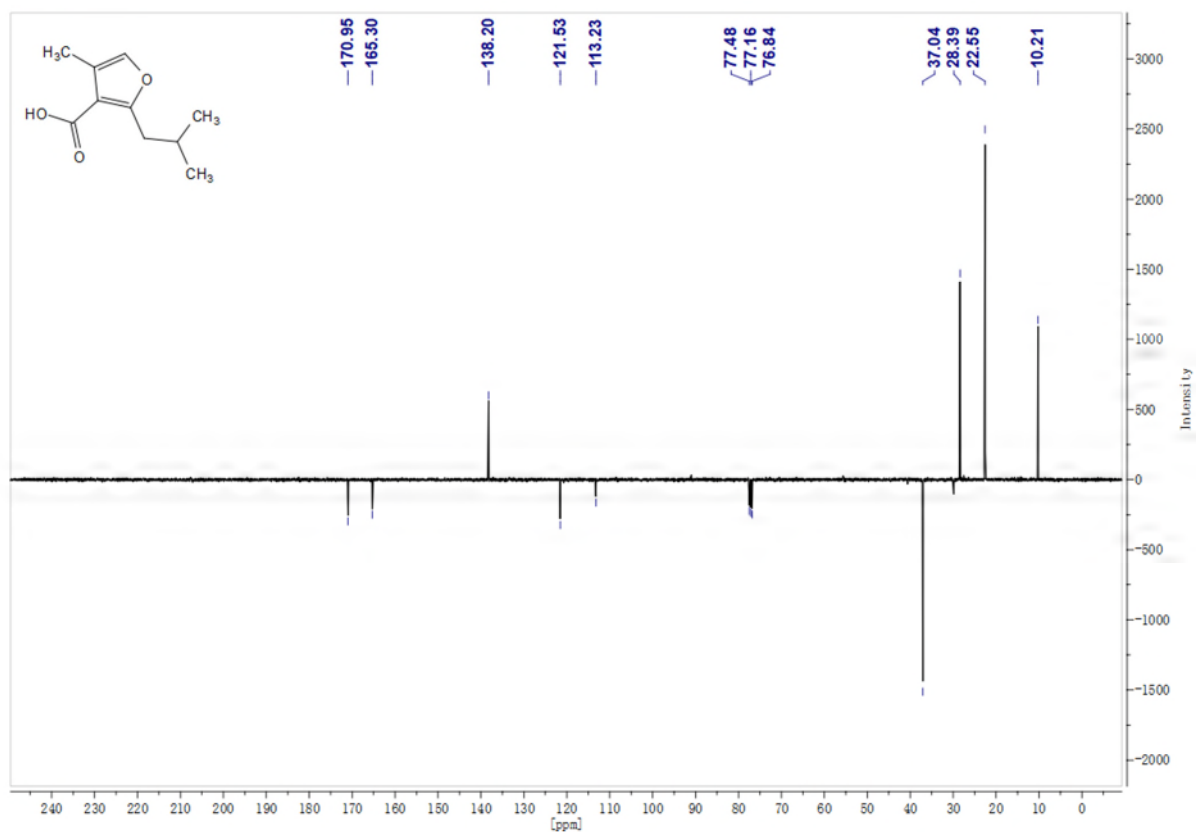
Supplementary Fig. 27 | <sup>13</sup>C-NMR spectrum of analogue 9.



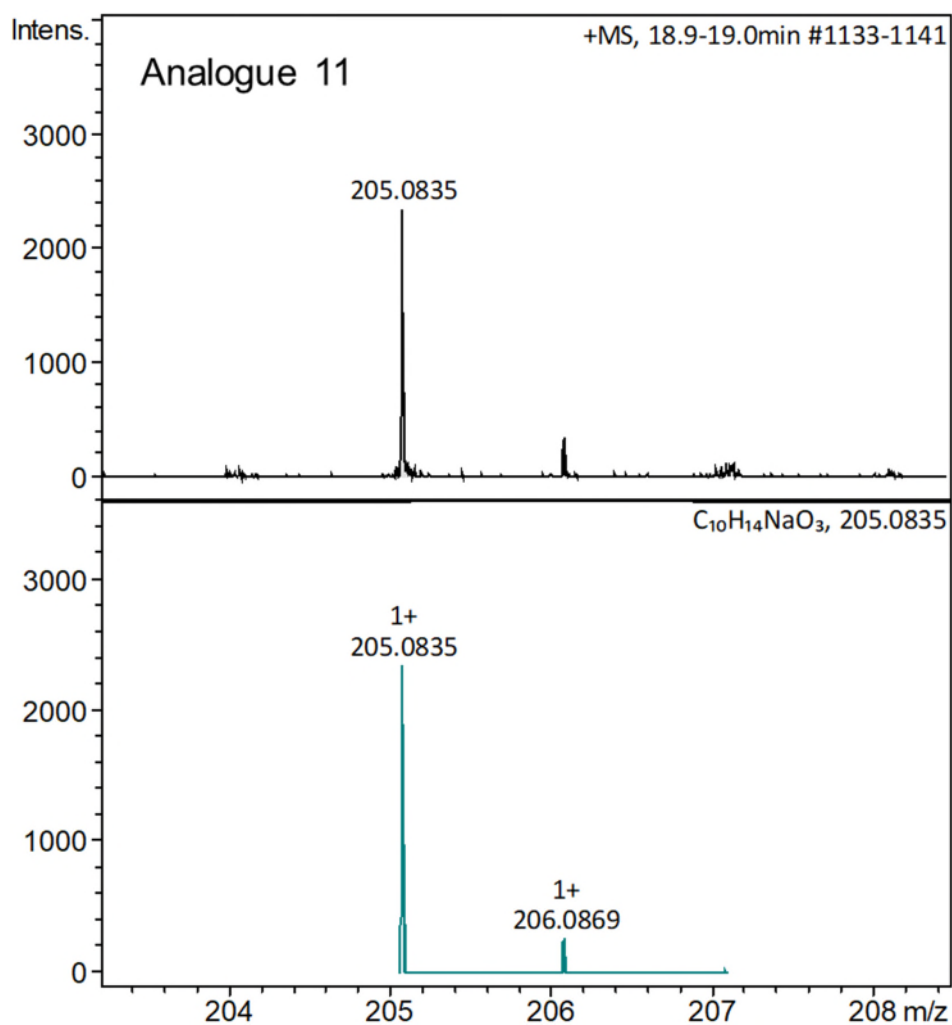
**Supplementary Fig. 28 | Comparison of measured (top panel) and simulated (bottom panel) mass spectra for analogue 9.**



Supplementary Fig. 29 | <sup>1</sup>H-NMR spectrum of analogue 11.

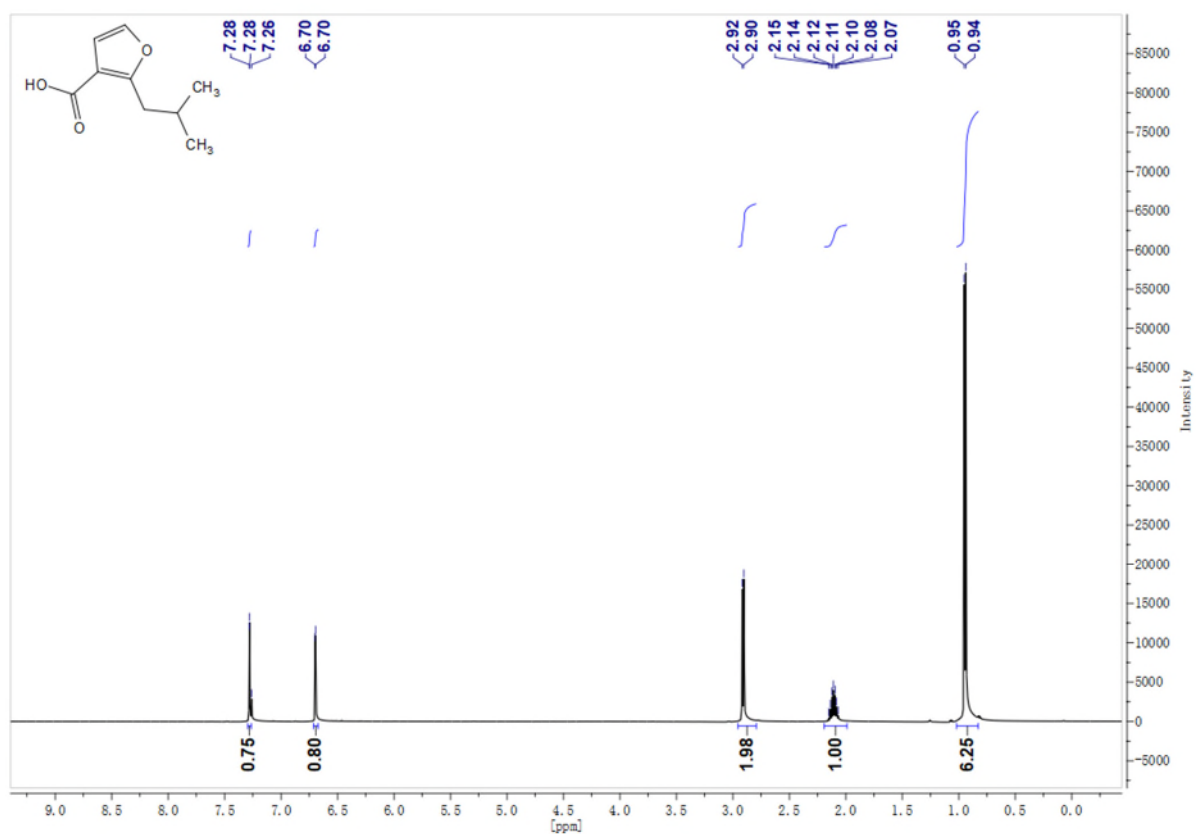


Supplementary Fig. 30 | <sup>13</sup>C-NMR spectrum of analogue 11.

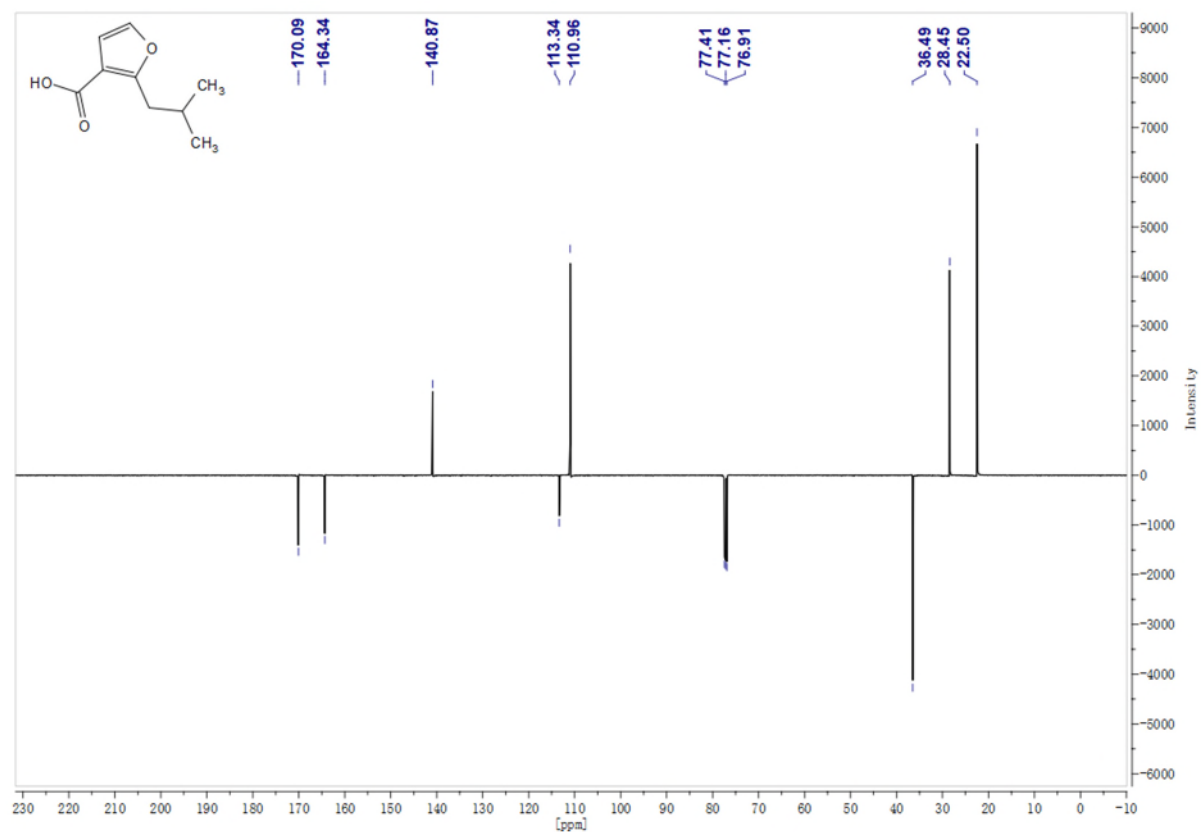


**Supplementary Fig. 31 | Comparison of measured (top panel) and simulated (bottom panel) mass spectra for analogue 11.**

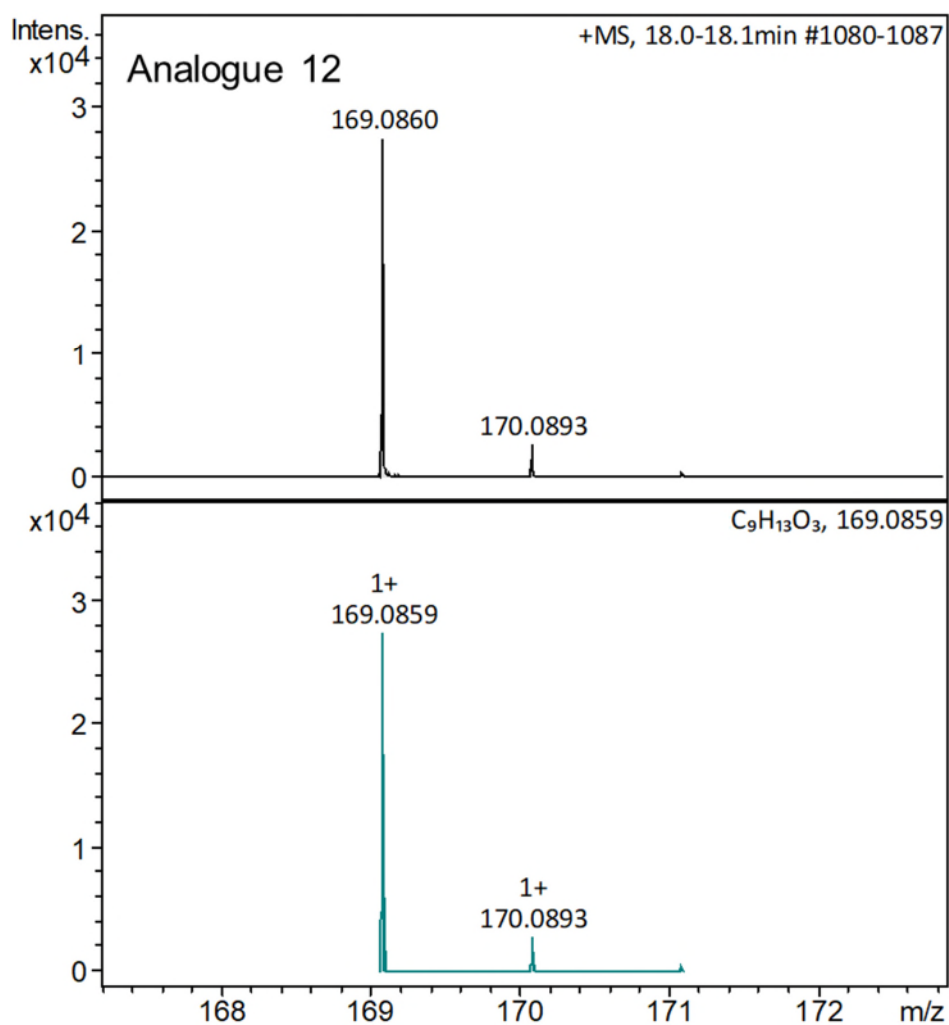




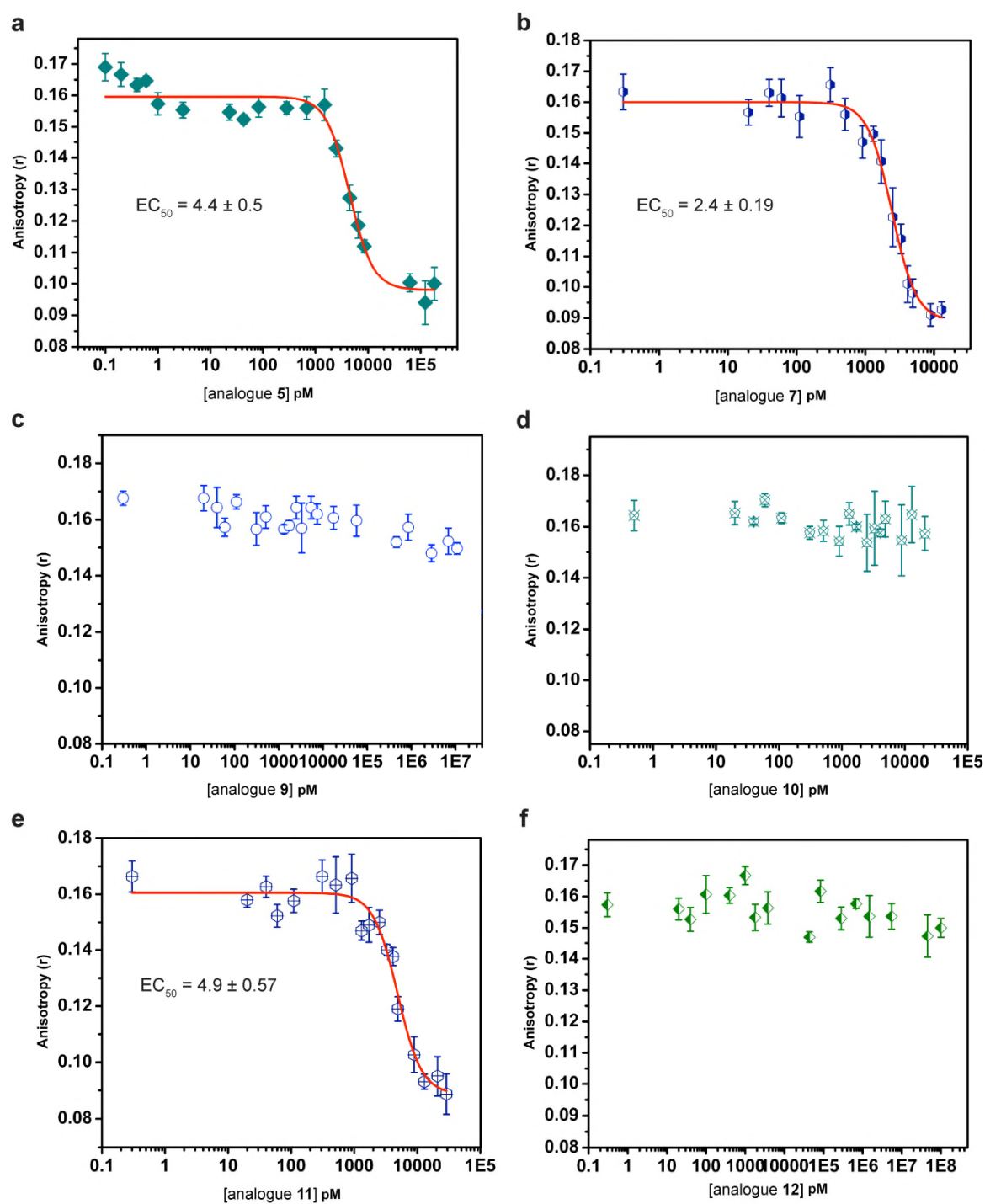
Supplementary Fig. 32 | <sup>1</sup>H-NMR spectrum of analogue 12.



Supplementary Fig. 33 | <sup>13</sup>C-NMR spectrum of analogue 12.



**Supplementary Fig. 34 | Comparison of measured (top panel) and simulated (bottom panel) mass spectra for analogue 12.**



**Supplementary Fig. 35 | FA plots for release of MmfR from MARE1 by selected MMF analogues.**

Data points are the mean of three independent technical replicates ( $n=3$ ) and error bars represent  $\pm 1$  standard deviation. The  $EC_{50}$  (nM) calculated from each data set is shown.