

Biosynthesis of β -Methylarginine residue of Peptidyl Nucleoside Arginomycin in *Streptomyces arginensis* NRRL 15941

Jun Feng^[a,b], Jun Wu^[a], Jie Gao^[a], Zhigui Xia^[b], Zixin Deng^[a], and Xinyi He^{*[a]}

^[a]State Key Laboratory of Microbial Metabolism and School of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai 200030 (China)

^[b]National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention; Key Laboratory of Parasite and Vector Biology, MOH; WHO Collaborating Centre for Malaria, Schistosomiasis and Filariasis; Shanghai 200025 (China)

***For Correspondence:** Xinyi He

Associate professor,
State Key Laboratory of Microbial Metabolism
and School of Life Science and Biotechnology
Shanghai Jiao Tong University
Shanghai, 200030, P. R. China

Tel: (86)-2162932943-2131(EXT)

Fax: (86)-2162932418

E-mail: xyhe@sjtu.edu.cn

Conflicts of Interest: We have no Conflicts of Interest to declare.

Keywords: β -methyl arginine; nucleoside antibiotic; gene cluster; L-aspartic acid; Biosynthesis

Supporting Information

Table S1 Primers used in this study.

Primer code	Sequences of primers (5'-3')
SCR-For	GACAGCCTSSSCTTCMTGTGG
SCR-Rev	YTGGCTSMRGTCTGYGGGGG
Argscr-For	TCTCGCCGTCCTCAATCTCG
Argscr-Rev	GCGTAGGCAGTTCGGTGTCC
ArgM-For	AACATATGGTGAGCCTGACGCCACTGGGA
ArgM-Rev	GAATTCGCAAAGGCGCTTGAACCGCTC
ArgN-For	AACATATGGATGAGGGATTCGAGCTGCGGC
ArgN-Rev	GAATTCGCGACCTCCGCGAGGCGCTGTC
TyrB-For	AACATATGGTGTTTCAAAAAGTTGA
TyrB-Rev	GAATTCAGTCACAGGCAATAAGGC

Table S2 Deduced functions of ORFs in the arginomycin biosynthetic gene cluster.

ORF	Locus	Size (aa)	Putative function	Source organism	Identities/positives	Accession NO.
Orf-1	1-201	66	FG-GAP repeat family protein	<i>Verticillium albo-atrum</i> VaMs.102	70%/87%	XP_002999611
Orf-2	340-909	189	transcriptional regulator, XRE family	<i>Kribbella flavida</i> DSM 17836	71%/85%	YP_003379448
Orf-3	979-1677	232	putative azaleucine resistance protein AzlC	<i>Providencia rustigianii</i> DSM 4541	61%/81%	ZP_05971682
Orf-4	1674-2009	111	predicted protein	<i>Streptomyces sviveus</i> ATCC 29083	60%/71%	ZP_06919003
Orf-5	2035-2538	167	stress-like protein	<i>Streptomyces viridochromogenes</i> DSM 40736	85%/90%	ZP_05534201
ArgA	2662-3930	422	putative efflux protein	<i>Streptomyces griseochromogenes</i>	69%/78%	AAP03124
ArgB	4136-4525	129	ABC transporter ATP-binding protein	<i>Streptomyces scabiei</i>	59%/64%	YP_003487422
ArgC	4438-5523	361	LuxR family transcriptional regulator	<i>Streptomyces coelicolor</i> A3(2)	49%/68%	NP_626920
ArgD	5576-6139	187	BlsM, putative nucleoside 2-deoxyribosyltransferase	<i>Burkholderia oklahomensis</i> EO147	41%/52%	ZP_02358672
ArgE	6132-7337	326	BlsD, cytosylglucuronic acid synthase	<i>Streptomyces griseochromogenes</i>	38%/50%	AAP03118
ArgF	7288-8346	368	BlsE, radical SAM	<i>Streptomyces griseochromogenes</i>	69%/79%	AAP03119
ArgG	8350-9306	318	BlsF, hypothetical protein	<i>Streptomyces griseochromogenes</i>	34%/45%	AAP03120
ArgH	9303-10529	408	BlsH, putative cytosinine synthase	<i>Streptomyces griseochromogenes</i>	63%/75%	AAP03122
ArgI	10574-11821	415	BlsJ, major facilitator superfamily MFS_1	<i>Catenulispora acidiphila</i> DSM 44928	40%/60%	YP_003117005
ArgJ	11909-13168	419	BlsI, putative ligase	<i>Streptomyces griseochromogenes</i>	48%/61%	AAP03123
ArgK	13249-15117	622	BlsK, hypothetical protein	<i>Catenulispora acidiphila</i> DSM 44928	50%/65%	YP_003116995
ArgL	15496-16134	212	BlsL, putativeleucyldemethylblastocidin S guanidino methyltransferase	<i>Streptomyces griseochromogenes</i>	45%/63%	AAP03126
ArgM	16394-17503	369	aminotransferase, class I and II	<i>Pseudomonas syringae</i> pv. <i>syringae</i> B728a	35%/51%	YP_233229
ArgN	17500-18513	337	methyltransferase	<i>Streptomyces</i> sp. <i>Mgl</i>	50%/65%	ZP_05000698
Orf+1	19557-19970	137	conserved hypothetical protein	<i>Streptomyces pristinaespiralis</i> ATCC 25486	90%/92%	ZP_06913994
Orf+2	19967-20251	94	conserved hypothetical protein	<i>Streptomyces sviveus</i> ATCC 29083	93%/96%	ZP_06919771

Orf+3	20302-20586	241	integrase	<i>Streptomyces sp. Mgl</i>	84%/96%	ZP_04999893
Orf+4	21660-22184	176	two-component system response regulator	<i>Streptomyces viridochromogenes</i> DSM 40736	93%/98%	ZP_07306497
Orf+5	22323-22525	400	arginyl tRNA synthetase	<i>Streptomyces viridochromogenes</i> DSM 40736	73%/75%	ZP_05534225
Orf+6	23437-24933	498	diaminopimelate decarboxylase	<i>Streptomyces viridochromogenes</i> DSM 40736	95%/98%	ZP_05534226
Orf+7	25115-26425	436	homoserine dehydrogenase	<i>Streptomyces viridochromogenes</i> DSM40736	98%/99%	ZP_05534227
Orf+8	26432-27490	352	threonine synthase	<i>Streptomyces ghanaensis</i> ATCC 14672	96%/99%	ZP_04685853
Orf+9	27677-28714	345	homoserine kinase	<i>Streptomyces viridochromogenes</i> DSM 40736	95%/97%	ZP_05534229
Orf+10	28775-30199	636	transcription termination factor Rho	<i>Streptomyces coelicolor</i> A3(2)	79%/82%	NP_733662.1

Table S3 Strains and plasmids used in this study.

Strains/plasmids	Relevant phenotype	Source
Strains		
<i>Streptomyces arginensis</i> NRRL 15941	Wild type producer of arginomycin	This study
<i>E. coli</i> DH10B	F- mcrA Δ (mrr-hsdRMS-mcrBC) ϕ 80dlacZ Δ M15 Δ lacX74 deoR recA1 endA1 ara Δ 139 D (ara,leu)1697 galU galK λ - rspL nupG	GIBCO BRL
<i>E. coli</i> BW25113	Δ (araD-araB)567, Δ lacZ4787(::rrnB-4), lacIp-4000(lacIQ), λ -, rpoS369(Am), rph-1, Δ (rhaD-rhaB)568, hsdR514	1
<i>E. coli</i> ET12567/pUZ8002	recF, dam-, dcm-, hsdS, Cml ^R , Kan ^R	2
EPI300-T1R	endA1 araD139 (ara, leu) 7697 galU galK-rpsL nupG trfA tonA dhfr	Epicentre Biotechnologies
<i>E. coli</i> BL21(DE3)pLysS	F- ompT hsdSB(rB-mB-) gal dcm (DE3) pLysS (Cam ^R)	This study
FJ1	<i>S. coelicolor</i> derivative harbouring the whole arginomycin gene cluster	This study
Plasmids		
pBluescript SK+	bla, lacZ, pUC origin, general clone vector	Stragagene
pSET152	aac (3) IV int (ϕ C31) oriT ori (PUC18)	3
pET-28a	KanR, pBR322 origin, T7 promoter, general protein expression vector	Novagen
PCT	pET-28a devriate, XbaI and SpeI were added	Chen <i>et.al.</i> unpublished
3F10	Cosmid harboring the biosynthetic gene cluster of arginomycin	This study

pFJ1	28.6 kb fragment of 3F10 digested with XbaI and HpaI ligation with pSET152	This study
pFJ2	ArgM expression vector, PCR product for ArgM ligated with pET-28a(+) with NdeI and EcoRI restriction sites	This study
pFJ3	ArgN expression vector, PCR product for ArgN ligated with pET-28a(+) with NdeI and EcoRI restriction sites	This study
pFJ4	TyrB expression vector, PCR product for TyrB ligated with pET-28a(+) with NdeI and EcoRI restriction sites	This study
pFJ5	<i>argM</i> ligated with PCT with NdeI and EcoRI restriction sites	This study
pFJ6	<i>argN</i> ligated with PCT with NdeI and EcoRI restriction sites	This study
pFJ7	PCT-argM ligated with PCT-argN with XbaI and SpeI restriction sites	This study

Reference

1. Gust, B.; Challis, G. L.; Fowler, K.; Kieser, T.; Chater, K. F., PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A* **2003**, *100* (4), 1541-6.
2. Paget, M. S.; Chamberlin, L.; Atrih, A.; Foster, S. J.; Buttner, M. J., Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J Bacteriol* **1999**, *181* (1), 204-11.
3. Bierman, M.; Logan, R.; O'Brien, K.; Seno, E. T.; Rao, R. N.; Schonher, B. E., Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* **1992**, *116* (1), 43-9.

Table S4 Analysis of TyrB function in the reaction.

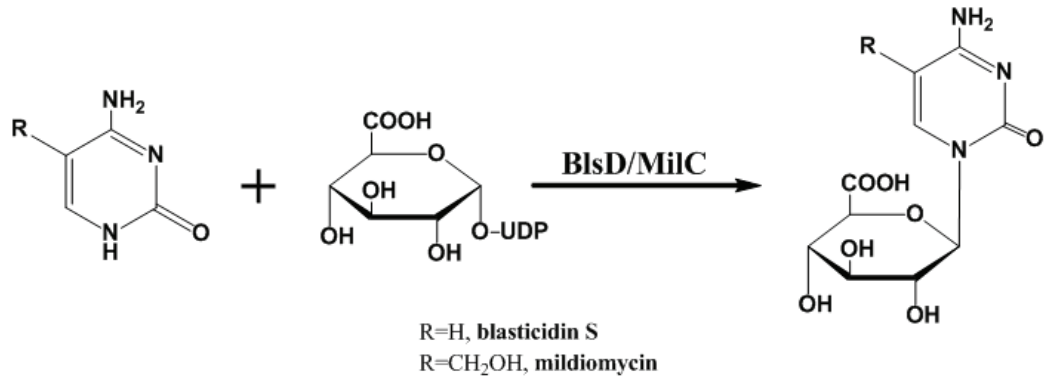
Enzyme ^[a]	Substrate	Activity ^[b]
ArgM	L-arginine	5-guanidino-2-oxopentanoic acid
ArgN	5-guanidino-2-oxopentanoic acid	5-guanidino-3-methyl-2-oxopentanoic acid
ArgM+ArgN	L-arginine	NA ^[c]
TyrB	L-arginine	NA
TyrB+ArgN	L-arginine	NA
TyrB+ArgM+ArgN	L-arginine	NA

[a]: This means the enzymes added in the reaction mixture.

[b]: The activity means the product generated in the reaction.

[c]: NA means no activity detected in the reaction mixture.

A)



B)



Figure S1 CGA synthase gene fragment was selected as the probe. A) Reaction catalyzed by BlsD/MilC. B) Alignment of DNA sequence of *blsD*, *milC* and other similarity genes, conserved sequences were indicated by black characters, the arrows indicate the locations of the degenerate primers are shown in SCF and SCR

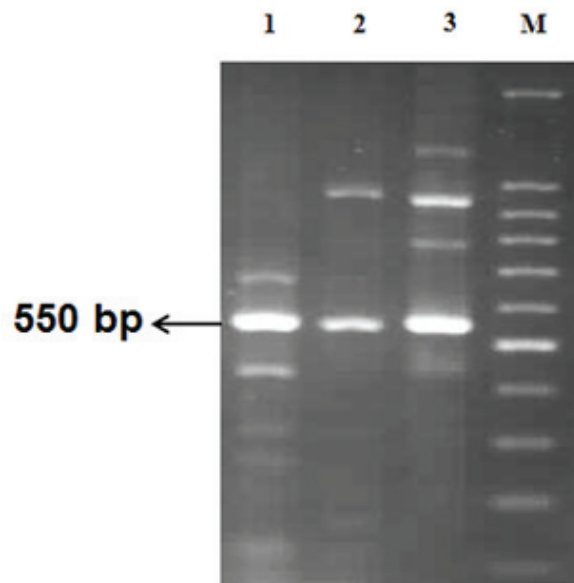


Figure S2 Amplification of putative CGA synthase gene fragment. 1 represent the dedicated PCR product from *Sv.rimofaciens* ZJU5119. 2 represent the dedicated PCR product from *S.griseochromogenes*. 3 represent the expected PCR product from *S. arginensis* NRRL 15941. The band of these expected sizes are indicated by the arrow.

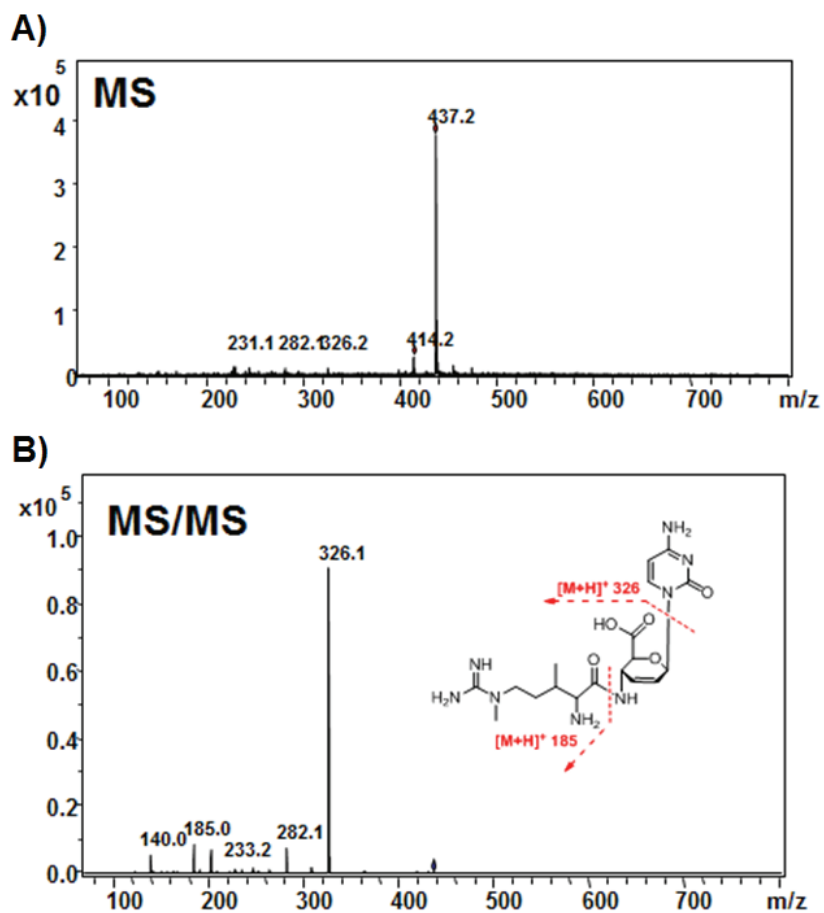


Figure S3 Heterologous production of arginomycin in *S. Coelicor* M145 was confirmed by MS analysis. A) MS ion of the peak with the retention time of 14.2 min from *S. coelicor* M145::pFJ1. 437.2 is the mass of protonated arginomycin; B) MS/MS fragmentation analysis of arginomycin in *S.coelicor* M145.

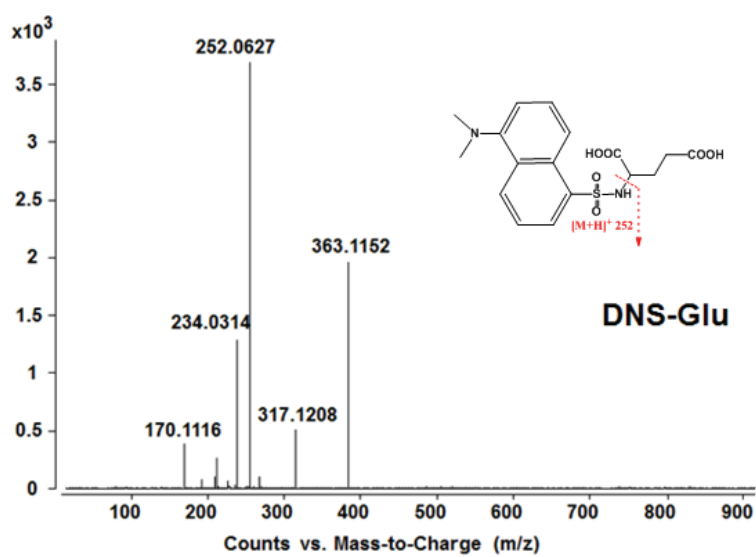
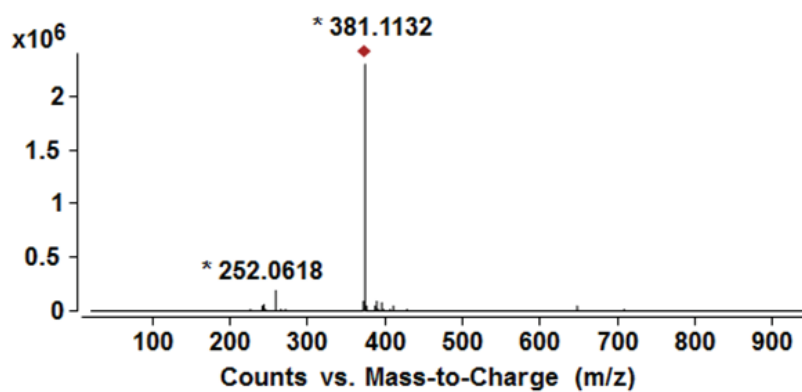


Figure S4 HR-ESI-MS analysis of DNS-Glu that is detected in the LC spectra. 381 represents the mass of protonated of DNS-Glu, 252 represents the fragment DNS-NH₂ of DNS-Glu.

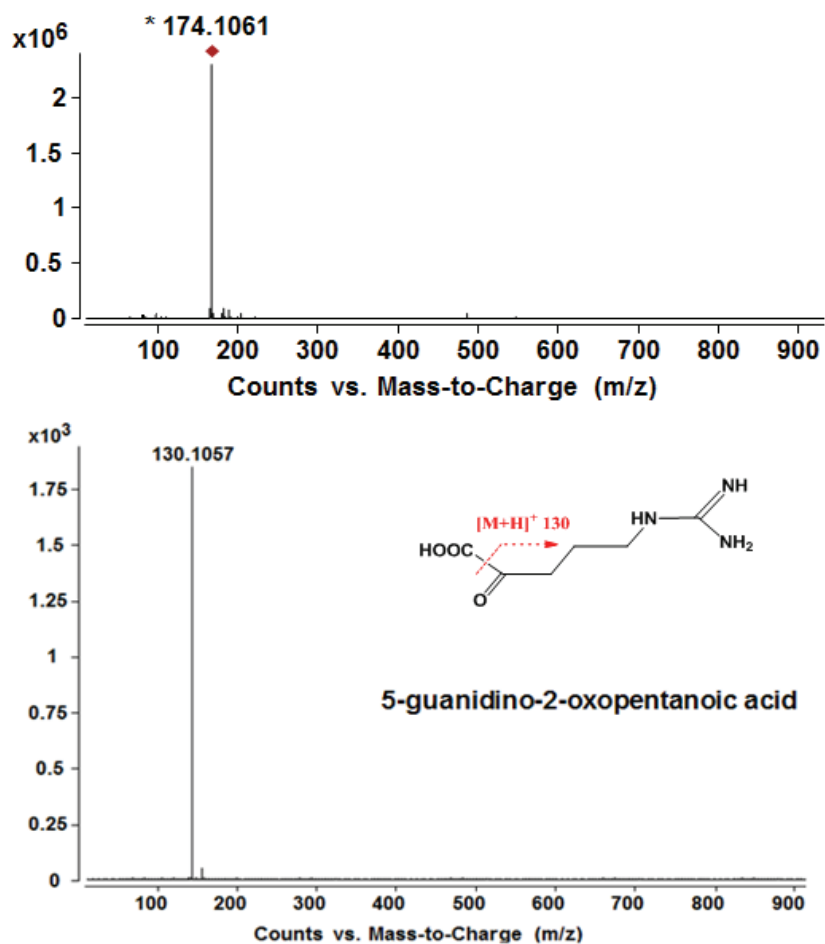


Figure S5 HR-ESI-MS analysis of 5-guanidino-2-oxopentanoic acid. 174 represents the mass of protonated of 5-guanidino-2-oxopentanoic acid, 130 represents the decarboxyted fragment of 5-guanidino-2-oxopentanoic acid.

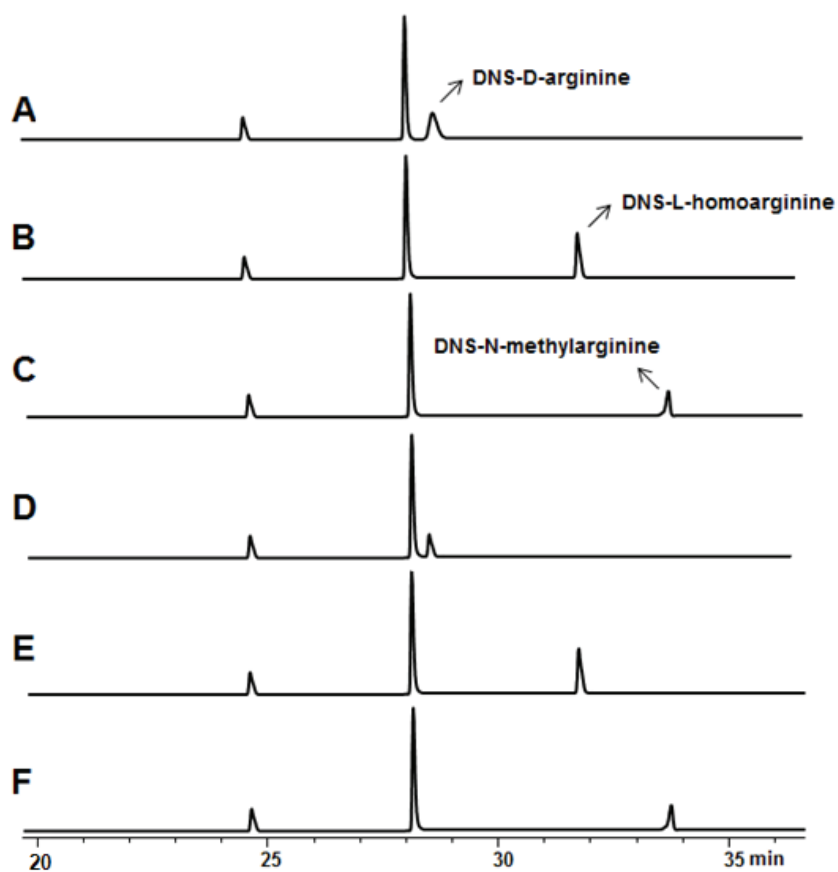


Figure S6 The substrate flexibility of ArgM. A, B and C is standard DNS-D-Arg, DNS-L-homoarginine and DNS-N-methylarginine, respectively. D, E and F is the *in vitro* ArgM assay with above D-arginine (D), L-homoarginine (E) and N-methylarginine (F) as substrate, respectively. No new DNS-Glutamate can be generated in the LC profile.

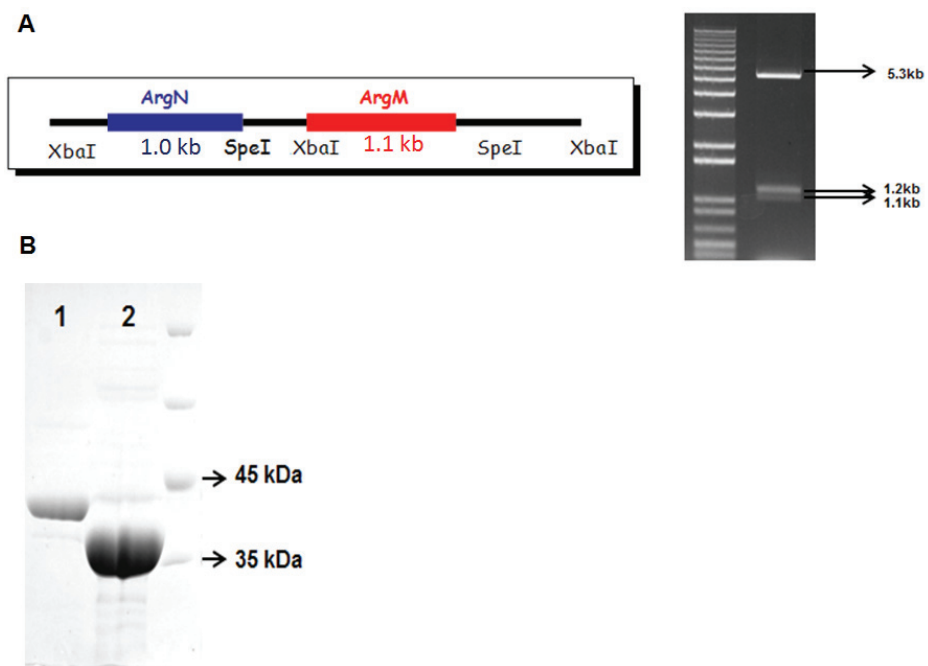


Figure S7 A) Construction and confirmation of vector pFJ7 which containing co-expressing ArgM and ArgN. Left panel is schematic representation of ArgM and ArgN coexpression vector. Right panel is enzymatic digestion by *NdeI/EcoRI* to confirm the construct. B) Lower panel is the SDS-PAGE of purified proteins. 1 represents purified ArgM, 2 represent purified ArgN.

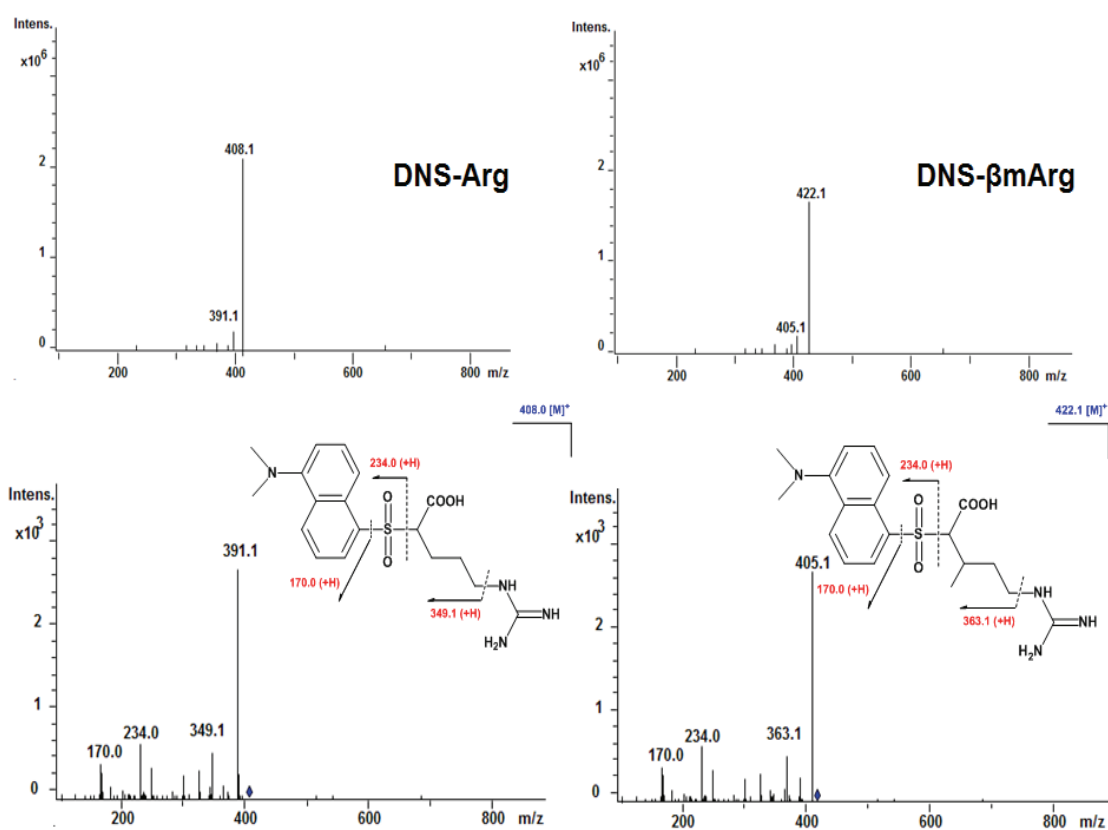


Figure S8 MS and MS/MS analysis of DNS-Arg and DNS-β-methyl Arg. Left panel in this figure is the MS data for the peak from Figure 6A upper panel. Right panel in this figure is the MS data for the peak from Figure 6A lower panel.

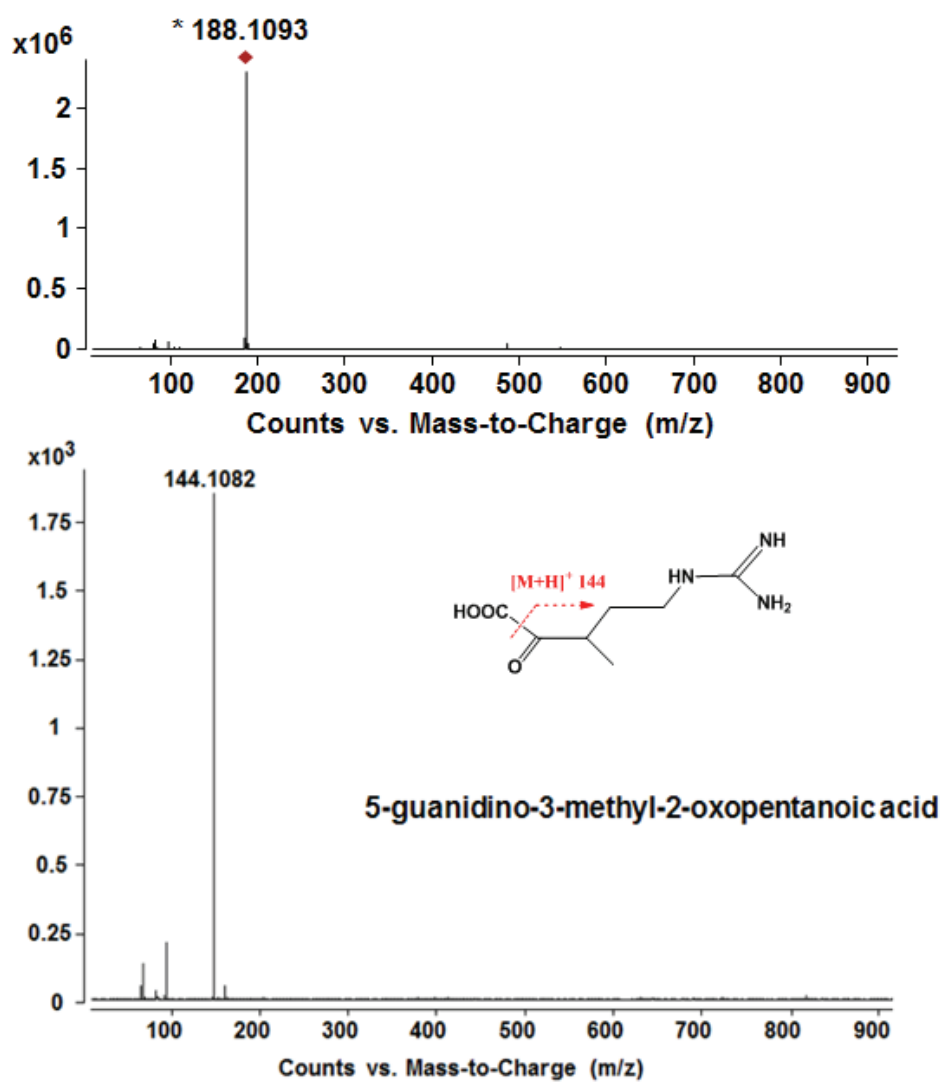


Figure S9 HR-ESI-MS analysis of 5-guanidino-3-methyl-2-oxopentanoic acid. 188 represents the mass of protonated of 5-guanidino-3-methyl-2-oxopentanoic acid, 144 represents the decarboxylated fragment of 5-guanidino-3-methyl-2-oxopentanoic acid.

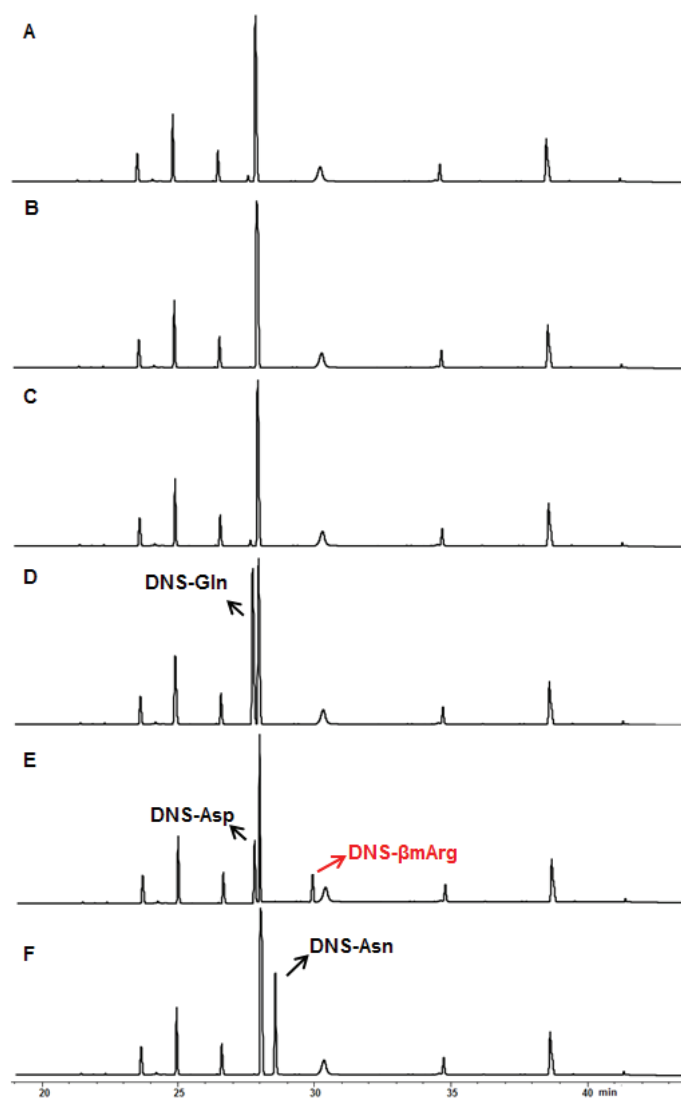


Figure S10 Survey of amino donors for the β -methylarginine formation. A, B and C represent inorganic amino donors $(\text{NH}_4)_2\text{SO}_4$, NH_4Cl and NH_4AC , respectively D, E and F represent organic amino donors L-Gln, L-Asp and L-Asn, respectively. DNS- β mArg was indicated in red character.

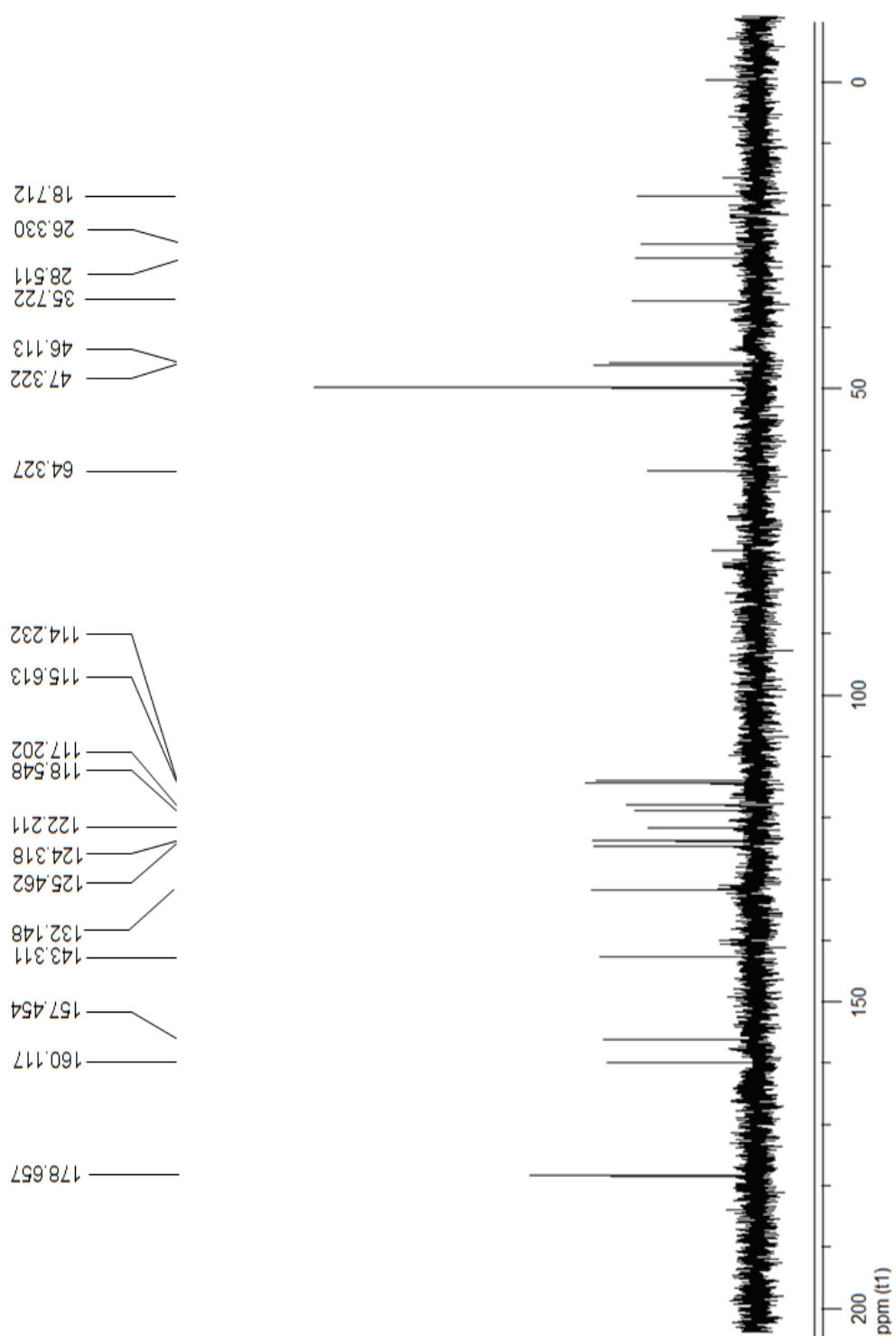


Figure S12 ^{13}C NMR (125 MHz) spectrum of DNS- β -methylarginine.

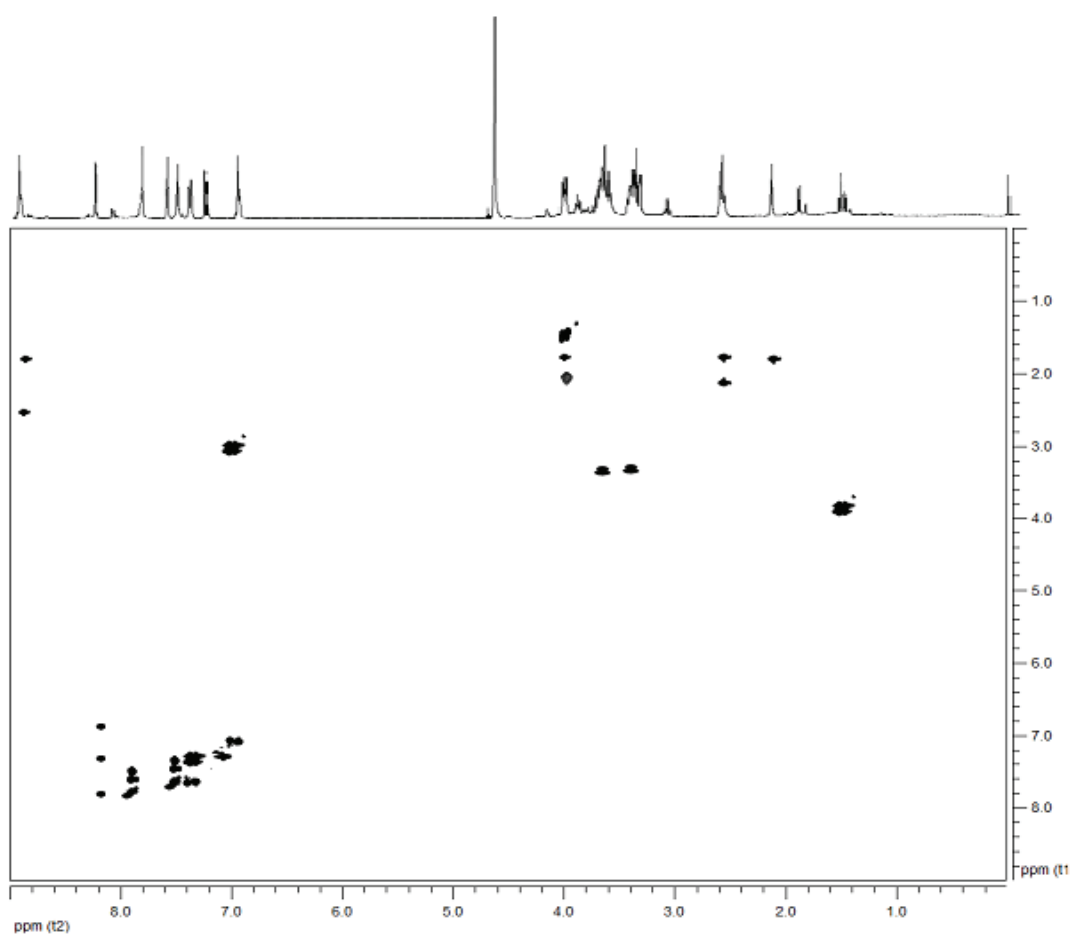


Figure S13 ^{13}C - ^1H COSY (500 MHz) spectrum of DNS- β -methylarginine.

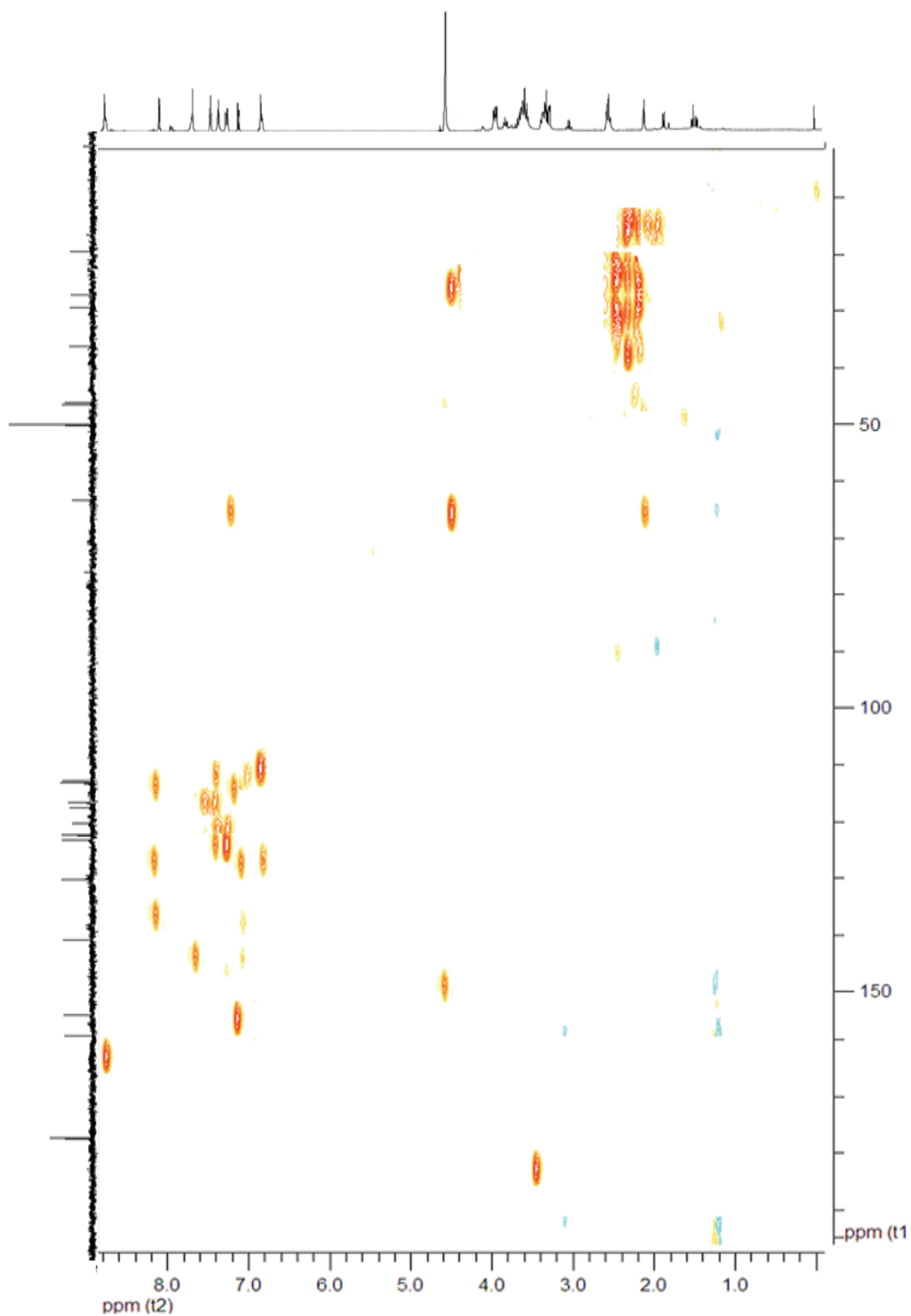


Figure S14 HSQC spectrum of DNS-β-methylarginine.