Electronic Supplementary Material (ESI) for RSC Advances.

Supplementary Information

From System Biology to System Chemistry: Metabolomic procedures enable insights into complex chemical reaction networks

Michaël Méret,[1†] Daniel Kopetzki,[2†] Thomas Degenkolbe,[1†] Sabrina Kleessen,[1] Zoran Nikoloski,[1] Verena Tellstroem,[3] Aiko Barsch,[3] Joachim Kopka,[1] Markus Antonietti,[2*] Lothar Willmitzer[1]

[1] Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, D-14476, Germany.

[2] Max Planck Institute of Colloids and Interfaces, Potsdam-Golm, D-14476, Germany.

[3] Bruker Daltonik GmbH, Bremen, D-28359, Germany.

* Corresponding author: email: markus.antonietti@mpikg.mpg.de

† contributed equally.

Content:

Methods

METHODS

Chemicals. Glycine, 1,2-$^{13}C_2$-glycine, 1-^{13}C-glycine, 2-^{13}C-glycine, ^{15}N-glycine, 2,2-D_2-glycine, glycolic acid, glyoxylic acid, oxamic acid, *N*-glycylglycine, iminodiacetic acid, glycineamide, sarcosine, alanine, serine, 3,6-dihydropyrazine-2,5-diol, and bis-*N*,*O*-trimethylsilyltrifluoroacetamide (BSTFA) were purchased (Sigma-Aldrich Chemie GmbH) at highest available purity. Millipore water was used throughout (Millipore GmbH).

The hydrothermal reaction. The hydrothermal decomposition of glycine was performed using an X-Cube Flash™ (ThalesNano Inc.) continuous flow reactor comprising a back pressure regulator for operation at 50–180 bar. The reactor was connected to a GX-271 auto-sampler (Gilson Inc.) for reproducible, automated sample collection. The flow-through capillaries of the reactor were made of corrosion and hydrothermal resistant Hastelloy C-22. Temperature and pressure of the reactor were preset to 180 °C at 100 bar or to 250 °C at 100 bar using pure water. 5 mL of 1% (w/v) solutions in water of isotope labeled or non-labeled glycine were injected through an injection loop and 15 mL were sampled after the reaction at ambient pressure and temperature. The reaction times were varied between 0.40 and 7.16 min by modulating the flow rate q of the continuous flow reactor with volume V. Residence times t_r in the flow reactor were calculated taking into account the density ratio of pure water [19] under reaction conditions, $\rho(T_r, p_r)$, and under ambient conditions, $\rho(T°,p°)$, according to Eq. (1).

$$t_r = \frac{V}{q} \cdot \frac{\rho(T_r, p_r)}{\rho(T°,p°)} \quad (1)$$

Series of 9–10 time points were generated. Three independent replicate reactions were performed at each time point. Reactions were stopped at ambient pressure and room temperature. Aliquots of the reaction products were stored frozen until further analysis to avoid the influence of slow reactions at room temperature.

Chemical profiling methods. ^{1}H-NMR spectra were recorded with a Bruker DPX-400 (400 MHz) spectrometer (Bruker BioSpin GmbH) in D_2O at 298 K. The chemical shift (δ) was determined relative to the HDO signal at 25 °C (δ 4.79 ppm). NMR data were processed by MestRe-C, a software package for desktop computing [20].

For gas chromatography-mass spectrometry (GC-MS) based chemical profiling, 16.7 µL of the glycine HTR product were rapidly dried in a VR Maxi vacuum concentrator (Jouan Nordic) and chemically derivatized by methoxyamination and trimethylsilylation [21-23] prior to analysis either by GC-electron impact ionization-time of flight-MS (GC-EI-TOF-MS) or by GC-atmospheric pressure chemical ionization-TOF-MS (GC-APCI-TOF-MS). Retention indices were calibrated by addition of a C_{10}, C_{12}, C_{15}, C_{18}, C_{19}, C_{22}, C_{28}, C_{32}, and C_{36} n-alkane mixture to each sample. Retention indices were used to align the paired GC-EI-TOF-MS and GC-APCI-TOF-MS runs with high accuracy as was described earlier [24].

GC-EI-TOF-MS chromatograms were acquired by a Pegasus III TOF fast scanning mass spectrometer (LECO Instrumente GmbH) set to a 70–600 *m/z* mass range, 20 scans s^{-1} acquisition rate and coupled to standard electron impact ionization at

−70 eV. GC was performed using an Agilent 6890N gas chromatograph with a CTC Combi PAL auto-sampler and PAL cycle composer software v1.5.0 (CTC Analytics AG). A VF-5ms capillary column with 30 m length, 0.25 mm inner diameter, 0.25 µm film thickness, and a 10 m EZ-guard pre-column (Varian Inc.) was operated with helium 5.0 carrier gas at 0.6 mL min^{-1}

constant flow starting with a 1 min hold at 70 °C, followed by a 9 °C min^{-1} ramp to 350 °C and a 5 min hold at final temperature. Sample volumes of 1µL were injected at 230 °C in splitless mode. GC-EI-TOF-MS chromatograms were acquired, baseline corrected, and exported in NetCDF file format using ChromaTOF software (Version 4.22; LECO). GC-MS data processing into a standardized numerical data matrix and compound identification by retention index and electron impact mass fragmentation were performed using TagFinder software [25]. Cases of co-elution which frequently occur in complex mixtures were solved by GC-EI-TOF-MS profiling with altered column polarity, *i.e.*, with a MDN-35 capillary column, 30 m length, 0.32 mm inner diameter, and 0.25 µm film thickness operated with 2 min isothermal time at 80 °C followed by a 15 °C min^{-1} ramp to 330 °C and a final hold of 6 min [23].

GC-APCI-TOF-MS profiles were generated using a maXis impact high mass accuracy mass spectrometer (Bruker Daltonics GmbH). The system was operated in a 20–800 *m/z* mass range at 4 scans s^{-1} acquisition rate and coupled to APCI operated in positive mode using nitrogen sheath gas. GC was performed using an Agilent 7890A system with a CTC PAL auto-sampler. A HP-5MS capillary column of 30 m, 0.25 mm inner diameter, and 0.25 µm film thickness was operated with 1.5 mL min^{-1} helium flow starting with a 1 min hold at 70 °C, followed by a 9 °C min^{-1} ramp to 350 °C and a 5 min hold at final temperature. Sample volumes of 1µL were injected at 230 °C with a 1/20 split ratio and in splitless mode. GC-APCI-TOF-MS chromatograms were acquired and processed by Bruker acquisition and DataAnalysis 4.1 software (Bruker Daltonics GmbH).

Synthesis of reference compounds

Potassium aminomalonate [1] and *N*-carboxyglycine [2] were synthesized and purified as described in the following. Diethylaminomalonate hydrochloride (5 g) was dissolved in 50 mL of 2 M

KOH. The solution was refluxed for 30 min and then chilled in an ice bath. Glacial acetic acid was added to neutralize the solution. After addition of 150 mL ethanol, a white solid precipitated. The product was recrystallized from 1:1 water: ethanol, yielding 2.65 g (71%) potassium aminomalonate (NMR in D_2O: δ_H 4.15, δ_C 59.2, 170.1). *N*-Carboxyglycine disodium salt was synthesized of 10.6 g Na_2CO_3 and 7.5 g glycine dissolved in 50 mL water. After filtration, 180 mL methanol was slowly added during 2 h while stirring. White crystals precipitated, which were filtered and washed with methanol and diethyl ether. Due to the equilibrium of product and free glycine in the presence of water, the material contained traces of glycine and was analyzed directly without further purification (NMR in D_2O: δ_H 3.59, δ_C 45.5, 163.2, 164.4, 179.4).

In analogy to metabolomic isotope labeling and tracing approaches, glycine of ambient isotope composition, 1,2-$^{13}C_2$-glycine, 1-^{13}C-glycine, 2-^{13}C-glycine, ^{15}N-glycine, and 2,2-D_2-glycine were subjected to HTRs. Isotope labeled reaction products allowed the mass spectrometric validation of the inferred structures of glycine HTR products without purification of the products from the complex reaction mixtures. Initial trials demonstrated that deuterium at the C2-position of glycine exchanged rapidly with H-atoms of water under the employed reaction conditions (data not shown). For this reason we neither used deuterium- nor ^{18}O-labeled glycine which is also likely to exchange a ^{18}O-atom with a ^{16}O-atom of water.

Glycine decomposition was monitored after freeze drying of the HTR product by GC-TOF-MS using mass fragment *m/z* = 102 of the *N*-(trimethylsilyl)-trimethylsilyl glycine ester (cf. the following paragraphs and Supplementary Table 1).

State-of-the-art methods for the chemical profiling and the characterization of compounds in complex mixtures

Characterization and identification of reaction products in complex mixtures

Metabolomics approaches solve the challenge of compound identification in complex mixtures. Due to time and abundance constraints, compounds typically cannot be purified from complex mixtures in sufficient amounts for conventional structure elucidation approaches. Following the parsimony principle, chemical characterization employed in metabolomics studies starts with obtaining a so-called chemical tag of each compound. Chemical tags are those detectable properties of compound, *e.g.*, specific mass fragments or fragmentation patterns, high resolution mass signals, chemical shifts or chromatographic retention information, which allow for selective, robust, and repeatable analysis within complex mixtures without required knowledge of each chemical structure [3,4].

We used the reproducible fragmentation patterns at nominal mass accuracy of GC-EI-TOF-MS to retrieve chemical tag information. We isolated the most abundant, specific, and selective, the so-called unique mass fragments and respective retention indices of each observed glycine HTR product [5]. This information was generated by automated mass spectral deconvolution using ChromaTOF software (Version 4.22; LECO). The deconvoluted mass spectra and their unique masses were manually validated with support by TagFinder software [6]. The unique mass fragments of this study are listed within supplementary tables (Supplementary Table 1, Supplementary Table 2). Retention time alignment of chromatograms was performed with the R package TargetSearch [7].

To obtain a hypothesis on the structure of observed compounds we analyzed each respective complete EI-MS spectrum manually. In addition, we used conventional mass spectral and retention index matching to pre-existing spectral libraries, such as the NIST mass spectral collection (http://chemdata.nist.gov/mass-spc/ms-search/) and the Golm Metabolome Database [8] and the substructure prediction tool of the Golm Metabolome Database [9]. The resulting hypothetical structures were cross-checked and if necessary revised using the EI-fragmentation mass spectra of 1,2-$^{13}C_2$-labeled, 1-^{13}C-labeled, 2-^{13}C-labeled, and ^{15}N-labeled glycine HTR products. The EI-MS fragmentation patterns are compiled within Additional data file S1 and can

be retrieved from the Golm Metabolome Database (http://gmd.mpimp-golm.mpg.de/) using the identifier codes provided by Supplementary Table 1.

The expected molecular formula of each inferred glycine HTR product was verified by GC-APCI-TOF-MS. This technology provided exact masses and mass isotopomer patterns of $[M+H]^+$ ions of glycine and of glycine HTR products. This information allowed for the deduction of lists of the most likely molecular formula of each observed component using SmartFormula software (Bruker Daltonics GmbH). The SmartFormula software calculates possible elemental compositions within a given mass tolerance window (mDa, ppm). All predicted molecular formulas are ranked according to the match of measured versus theoretical isotopic pattern (mSigma), cf. Table S1. The top ranking prediction was validated by the predicted molecular formula of the respective ^{13}C- and ^{15}N-labelled $[M+H]^+$ ions (Tab. S1). This process enabled the unambiguous generation of consensus molecular formulas.

In agreement with metabolomic procedures only those structures which were deemed essential for the interpretation of reversed engineered reaction networks of glycine HTR were identified. For this process authenticated commercially available or synthesized reference substances (Supplementary Table 1) were used. To accept a compound identification by GC-MS, a full mass spectral match and retention index deviation < 1.0% of the co-analyzed pure reference substance was required[10].

Positional labeling analysis

The EI-TOF-MS fragmentation patterns of trimethylsilylated sarcosine, alanine and serine were used to confirm the origin of the C-atom added to the glycine backbone of each of these compounds. The trimethylsilylated substance generated in each case at least one EI-MS fragment that contained C2 and C3 of the above compounds [11,12]. The mass shift observed in products obtained by HTR of 1-^{13}C-labeled, 2-^{13}C-labeled, and ^{15}N-labeled glycine compared to the products of non-labeled glycine proved the origin of the added C-atom.

In detail, the *N*,*O*-ditrimethylsilylserine trimethylsilyl ester generated the EI-MS fragment *m/z* = 204 which contains C2 and C3 (Supplementary Fig. 3A). The 1-^{13}C-glycine product showed no mass shift of this fragment whereas the 2-^{13}C-glycine product had a mass shift of +2. Therefore, C3 of serine originated from the C2-position of glycine. Likewise, the *N*-(trimethylsilyl)-,

trimethylsilyl esters of alanine (Supplementary Fig. 3B) and sarcosine (Supplementary Fig. 3C) generated the EI-MS fragment *m/z* = 116 which contains C2 and C3 of these compounds. ^{15}N-labeling resulted in a +1 mass shift confirming the presence of the 2-amino moiety within these fragments. The 1-^{13}C-glycine products showed no mass shift of this fragment whereas the 2-^{13}C-glycine product had a mass shift of +2. Therefore, C3 of alanine and the *N*-methyl moiety of sarcosine originated from the C2-position of glycine. The respective M[-15]$^{+}$ mass fragments of serine (3TMS), alanine (2TMS), and sarcosine (2TMS), *m/z* = 306, 218, and 218, confirmed the expected mass shifts for the full carbon backbones of the respective structures (Supplementary Fig. 3).

State-of-the-art **computational methods for analysis of time-resolved metabolomics data**

Reconstruction of chemical reactions through which compounds are transformed spontaneously or in presence of catalysts still remains the most prominent problem in chemistry. Solutions to this problem will have applications in other scientific disciplines where the mechanisms of operations leading to system-wide behavior are to be elicited from high-throughput data (*e.g.*, systems biology). Chemical reactions are fully determined by their: *stoichiometry*, *i.e.*, the number of molecules of substrates and products consumed and produced by the reaction, respectively, while ensuring mass preservation, *thermodynamics, i.e.*, activation energy, and *kinetics* accounting for reactant concentration, surface area available for reactants' contact, pressure, activation energy, temperature, and properties of the catalysts, if any.

Therefore, given a set of data, the formal requirement of any method for reconstructing chemical reactions consists of either fully specifying the abovementioned determinants or at least providing feasible ranges for their values. The data sets employed by the methods for resolving this inverse problem can be obtained from multiple *stationary* conditions, *i.e.*, steady states at which there is no change in reaction rates/ fluxes and concentrations of substances or *non-stationary* conditions, usually described by time series of varying length and sampling frequency, depending on the experimental set-up. Therefore, design, analysis, and testing of methods for reconstructing chemical reactions by employing (non-)stationary metabolomics or in this case systems chemistry data is a first necessary step towards resolving the problem of determining and discriminating between reaction mechanisms.

The simplest approach for network inference from (non-)stationary data sets relies on applying similarity measures. In addition, methods borrowed from statistical inference, *e.g.*, (non-)linear regression and its derivatives, Bayesian inference, and econometrics models, *e.g.*, Granger causality have been applied in this context. A recent review provides the most comprehensive comparative analysis of the ability of these approaches to reconstruct topology of the underlying reactions, therein, biochemical reactions specifying the interactions between genes [13]. These methods have been widely used in reconstruction of metabolic networks, composed of metabolites and enzymatic reactions facilitating their transformation [14]. However, all of these methods fail to meet the aforementioned implicit requirements largely due to the fact that they focus only on reconstructing the topology of pairwise relationships and, thereby, ignore the fact that these networks are hypergraphs, capturing many-to-many relationships, *e.g.*, bimolecular reactions [15].

Our contribution is two-fold: (1) Devising a novel method for reconstructing networks of chemical reactions solely from (non-)stationary metabolomics data based on interleaving the information contained in all available replicates, and (2) implementing and applying the method to discern patterns of interactions from glycine HTRs.

Computational method for the reversed engineering of reaction networks from (non-) stationary states

We assume that there are data capturing the levels of chemical compounds in non-stationary state from l different experiments from which measurement are obtained at T time points. In the case of stationary states, $T = 1$. Let the data for n compounds at time point t, $1 \leq t \leq T$, be represented by the matrix form ${X_{n \times l}}^{t}$, with rows representing substances and columns denoting the experiments. Let us also assume that there are m reactions through which the n substances are transformed and that the reaction rates obey mass-action law. Then, the rate of the i^{th} $(1 \leq i \leq m)$ reaction at time point t is given by:

$$v_{i,t} = k_i \prod_{j=1}^{n} {x_{j,t}}^{\alpha_{ij}}, \tag{2}$$

where k_i denotes the (temperature and pressure-dependent) kinetic constant, $x_{j,t}$ stands for the level of the j^{th} substance at time point t, and α_{ij} is the stoichiometric coefficient of the j^{th} substance appearing as a substrate in the i^{th} reaction. Let $v^*_{i,t}$ denote the reaction rate at another non-stationary state, *i.e.*, the state in which the network is at time point t starting from different initial levels/concentrations for the chemical substances assuming everything else constant; analogously to Eq. (2), we then have:

$$v^*_{i,t} = k_i \prod_{j=1}^{n} x^{*\,\alpha_{ij}}_{j,t}. \tag{3}$$

By dividing Eqs. (2) and (3), one obtains:

$$\frac{v_{i,t}}{v^*_{i,t}} = \prod_{j=1}^{n} \left(\frac{x_{j,t}}{x^*_{j,t}}\right)^{\alpha_{ij}}, \tag{4}$$

in which the effect of the kinetic constants is removed. By taking logarithms from the left- and right-hand side, one has the following:

$$\log \frac{v_{i,t}}{v^*_{i,t}} = \sum_{j=1}^{n} \alpha_{ij} \log \frac{x_{j,t}}{x^*_{j,t}}. \tag{5}$$

Introducing the variable substitution $z_{i,t} = \log \frac{v_{i,t}}{v^*_{i,t}}$ and $y_{j,t} = \log \frac{x_{j,t}}{x^*_{j,t}}$, and summing over the reaction rates of all reactions yield:

$$\sum_{i=1}^{m} z_{i,t} = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} \alpha_{ij} \right) y_{j,t}. \tag{6}$$

The only known variables are $y_{j,t}, (1 \le j \le n, 1 \le t \le T)$, which can be obtained from the matrix X, containing the experimental data, by using the transformation described in Eq. (5). Moreover, we can rewrite Eq. (6) in order to express the variable $y_{k,t}$, resulting in:

$$y_{k,t} = \frac{\sum_{i=1}^{m} z_{i,t}}{\sum_{i=1}^{m} \alpha_{ik}} - \sum_{j=1, j \neq k}^{n} \left(\frac{\sum_{i=1}^{m} \alpha_{ij}}{\sum_{i=1}^{m} \alpha_{ik}} \right) y_{j,t}. \quad (7)$$

If the vectors $y_{k,t}$ are standardized to be of unit norm (vector length), then Eq. (7) becomes:

$$y_{k,t} = - \sum_{j=1, j \neq k}^{n} \left(\frac{\sum_{i=1}^{m} \alpha_{ij}}{\sum_{i=1}^{m} \alpha_{ik}} \right) y_{j,t}. \quad (8)$$

The problem of reverse engineering m reactions from l metabolomics measurements (or equivalently $\frac{l(l-1)}{2}$ log-ratios of levels/concentrations) can be regarded as a regression problem — albeit on a modified set of data. Since networks of chemical reactions are known to be sparse, *i.e.*, the number of substrates transformed via a reaction is bounded from above by a small constant — 2 or 3, Eq. (8) has to be solved with a small number of non-zero entries, denoting the likeliest substrates participating in a reaction.

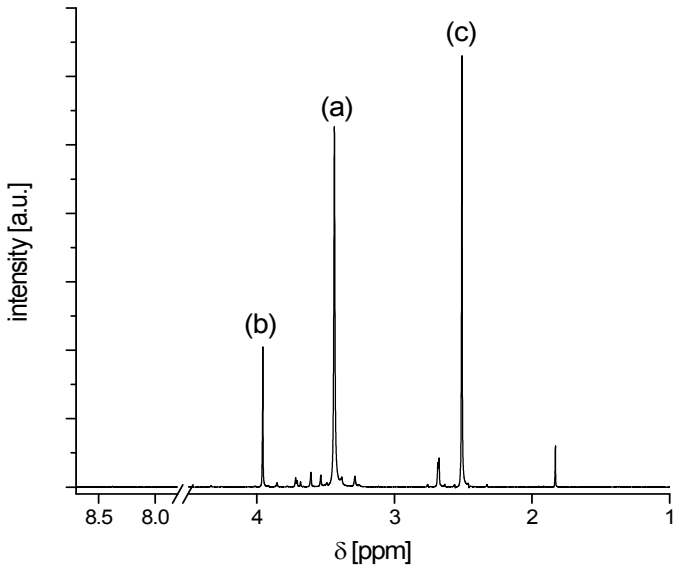
We applied LASSO regression in order to solve Eq. (8), since this method enjoys the favorable properties of both subset selection and ridge regression; namely, it yields a small number of predictors of high predictive power [16]. To this end, for each time point, we determined the ratios of compound's levels over the three available replicates, resulting in the $y_{k,t}$ vectors. Hence, for every time point, each compound was described with a vector containing 6 entries, termed *ratio replicates*, corresponding to the all possible ratios between the 3 replicates. Eq. (8) was solved with LASSO regression by using the R package *lars*. The coefficients of the LASSO regression were 3-fold cross validated in accordance with the 6 ratio replicates in order to produce robust estimates of the LASSO coefficients. The R^2 statistics was equal to one for all considered models. Compounds with regression coefficients of same sign can be interpreted, according to Eq. (5), to be on the same side of a reaction. Only compounds for which all replicates were measured at all time points were used in the analysis. These compounds were glycine (**1**), carbonic acid (**2**), glycolic acid (**3**),glyoxylic acid (**4**), iminodiacetic acid (**7**), sarcosine (**8**), *N*-carboxyglycine (**9**), glycineamide (**11**), glycine-*N*-methylamide (**12**), alanine

(**13**), serine (**14**), hydantoin (**16**), 3,6-dihydropyrazine-2,5-diol (**17**), pyrazine-2,5-diol (**18**), pyrazine-2,3,5-triol (**20**) of the glycine HTRs at both temperatures as well as *N*-glycylglycine (**6**) of the glycine HTRs at 180 °C and 3,6-dihydropyrazine-2,3,5-triol (**19**) of the glycine HTRs at 250 °C. For some of the time points no statistically robust relationships could be obtained, and, thus, these were excluded from further interpretations.

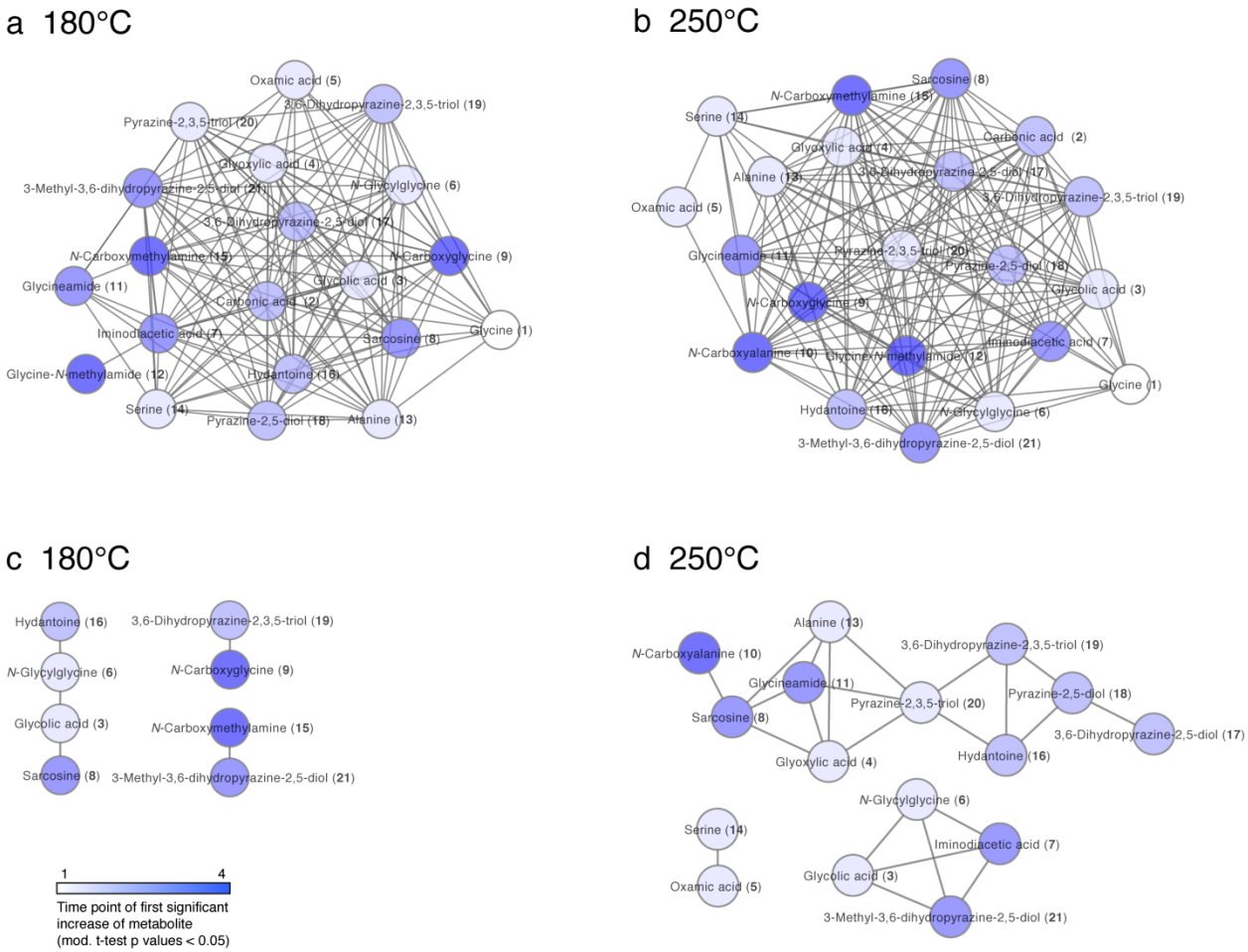
Descriptive and inferential statistical analyses and correlation-based network approaches

Data were visualized and analyzed by using the R software (http://www.r-project.org/) and the Bioconductor packages. Modified t-test statistics were calculated using the R package limma [17]. The time point of the first appearance of a glycine HTR product was defined at the first significant increase relative to t_0, $p < 0.05$. The original response matrix (Supplementary Table 2) was transformed into a Hamming matrix by encoding non-significant changes or decreases as 0 and significant increases as 1. K-medoids clustering was performed by the package fpc [18]. The number of clusters k was determined based on the Calinski-Harabasz index [19] with the function pamk of fpc. Spearman correlation coefficients and first-order partial Spearman correlation coefficients of metabolite pairs were calculated separately for each temperature condition after $\log_2$-transformation of means at each time point. Non-directed networks were constructed with nodes representing the compounds and edges representing significant Spearman correlation coefficients, $p < 0.05$. The significance threshold was determined using 10,000 randomized data sets where the time series data were shuffled independently for each metabolite. With $p < 0.05$, the Spearman correlation coefficients were $r > 0.636$ and 0.609, for 180 °C and 250 °C time series, respectively. The more conservative threshold $r > 0.636$ was applied for both correlation networks. With $p < 0.05$, the first-order partial Spearman correlation coefficients were $r > 0.585$ and 0.551, for 180 °C and 250 °C time series, respectively. The more conservative threshold $r > 0.585$ was applied for both partial correlation networks. Probabilistic principal component analysis was performed after $\log_2$-transformation of unit variance scaled and centered data by using the Bioconductor package pcaMethods [20].

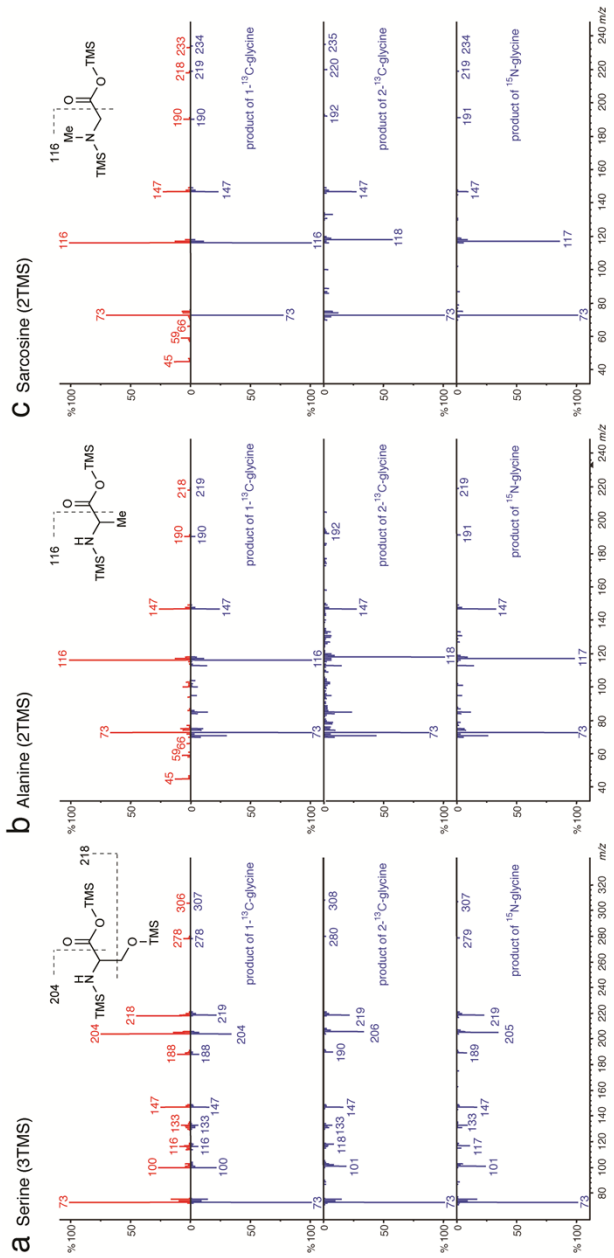
Supplementary Results



Supplementary Figure 1 ^{1}H-NMR profile of a 1% (w/v) glycine solution subjected to hydrothermal decomposition at 250 °C and 100 bar for 2.70 min. The main observed chemical shifts agreed with previous studies [21] and represent (a) glycine (**1**) δ = 3.35 (s, 2H, $H_2NC\mathit{H}_2COOH$), (b) 3,6-dihydropyrazine-2,5-diol (**17**) δ = 3.85 (s, 2H, $(HNC\mathit{H}_2CO)_2$) (c) methylamine δ = 2.42 (s, 3H, $H_2NC\mathit{H}_3$) glycine (**1**) (a). Several low intensity peaks were detectable between 1.5–4 ppm. The signals were difficult to match unambiguously to the known glycine decomposition products. At low field only a trace peak at δ = 8.40 was detectable.

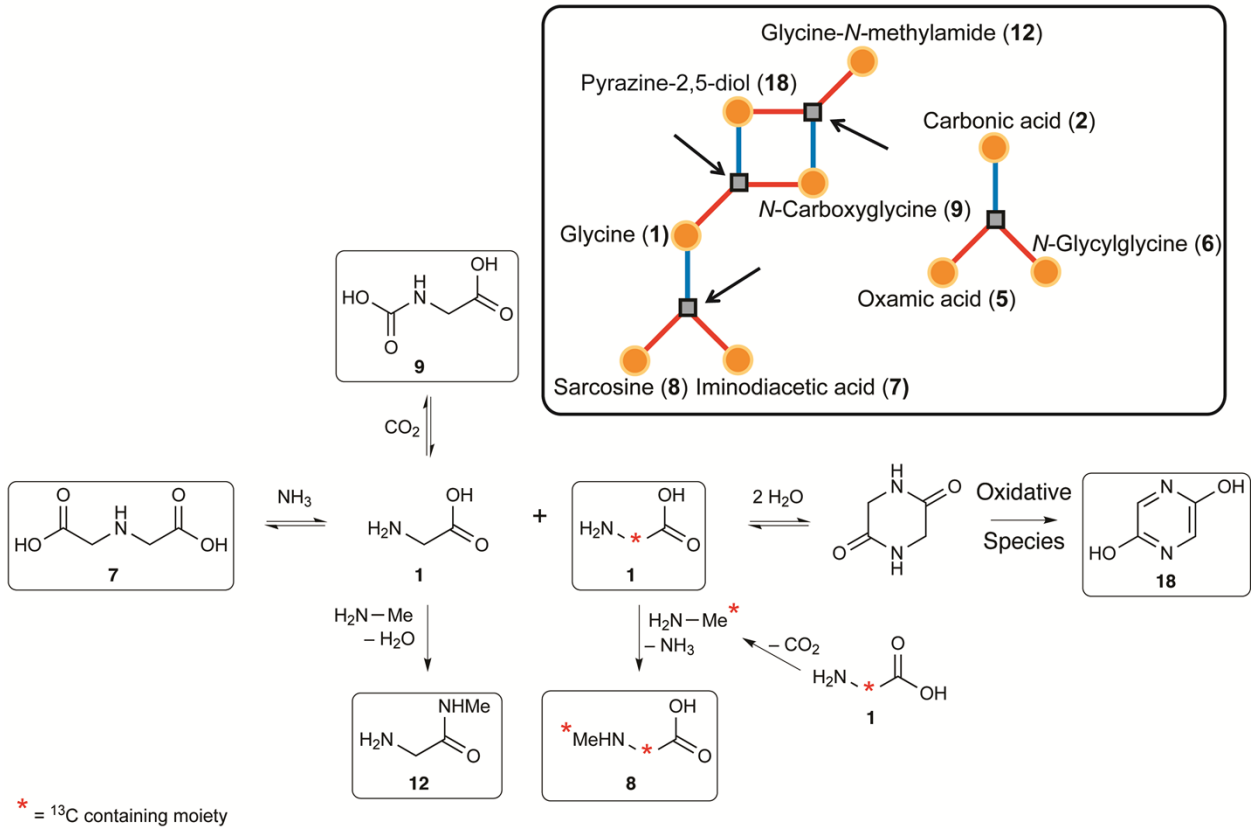


Supplementary Figure 2. Correlation networks of the time-resolved profiles of glycine HTRs at 180 °C and 250 °C. Edges denote the association of glycine HTR products with a significant correlation ($p < 0.05$). Blue color coding indicates the time of appearance of each compound as deduced by the Hamming matrix (Fig. 2B). Spearman correlation networks did not allow a differentiated analysis of associations between the products, because with 143 and 174 edges at 180 °C (**a**) and 250 °C (**b**), respectively, these networks were close to complete *i.e.*, a network containing all 210 edges corresponding to all pairs of metabolites. First-order partial Spearman correlation networks of glycine HTRs at 180 °C (**c**) and 250 °C (**d**) where spurious correlations were removed.

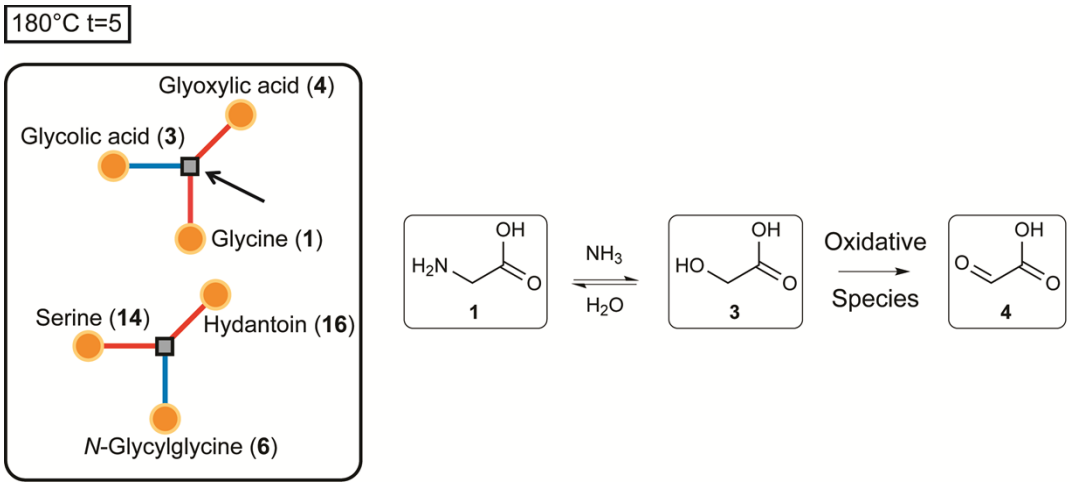


Supplementary Figure 3. Electron impact ionization time of flight mass spectra (EI-TOF-MS) at nominal mass resolution of trimethylsilylated products from glycine HTRs. The EI-TOF-MS of trimethylsilylated serine (**14**) (**a**), alanine (**13**) (**b**) and sarcosine (**8**) (**c**) were obtained by manually supervised automated spectral deconvolution from the chemical profiles of HTR products. The product of non-isotopic labeled glycine (top) is matched head to tail to the products of positional labeled 1-^{13}C-glycine, 2-^{13}C-glycine, and ^{15}N-glycine (from top to bottom). The structures and the mass to charge ratios (*m/z*) of the main EI-fragments are depicted

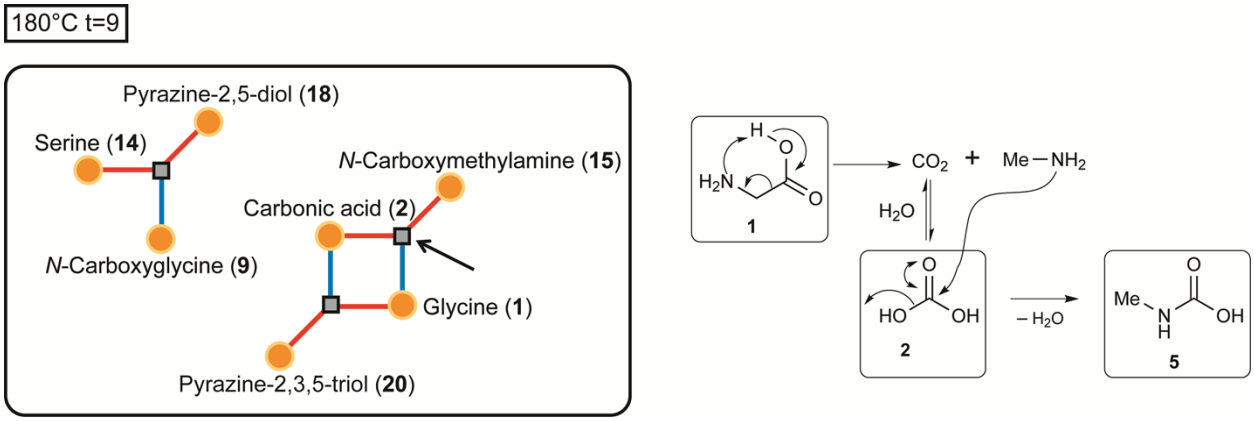
and indicated at the top of each subfigure. The isotope induced mass shifts of the fragments *m/z*=204 of serine and *m/z*=116 of alanine or sarcosine are in agreement with the conclusion that the C3-position of serine and alanine originate from the C2-position of glycine (**1**). Likewise, the *N*-methyl moiety of sarcosine (**8**) is demonstrated to also originate from the C2-position of glycine (**1**).



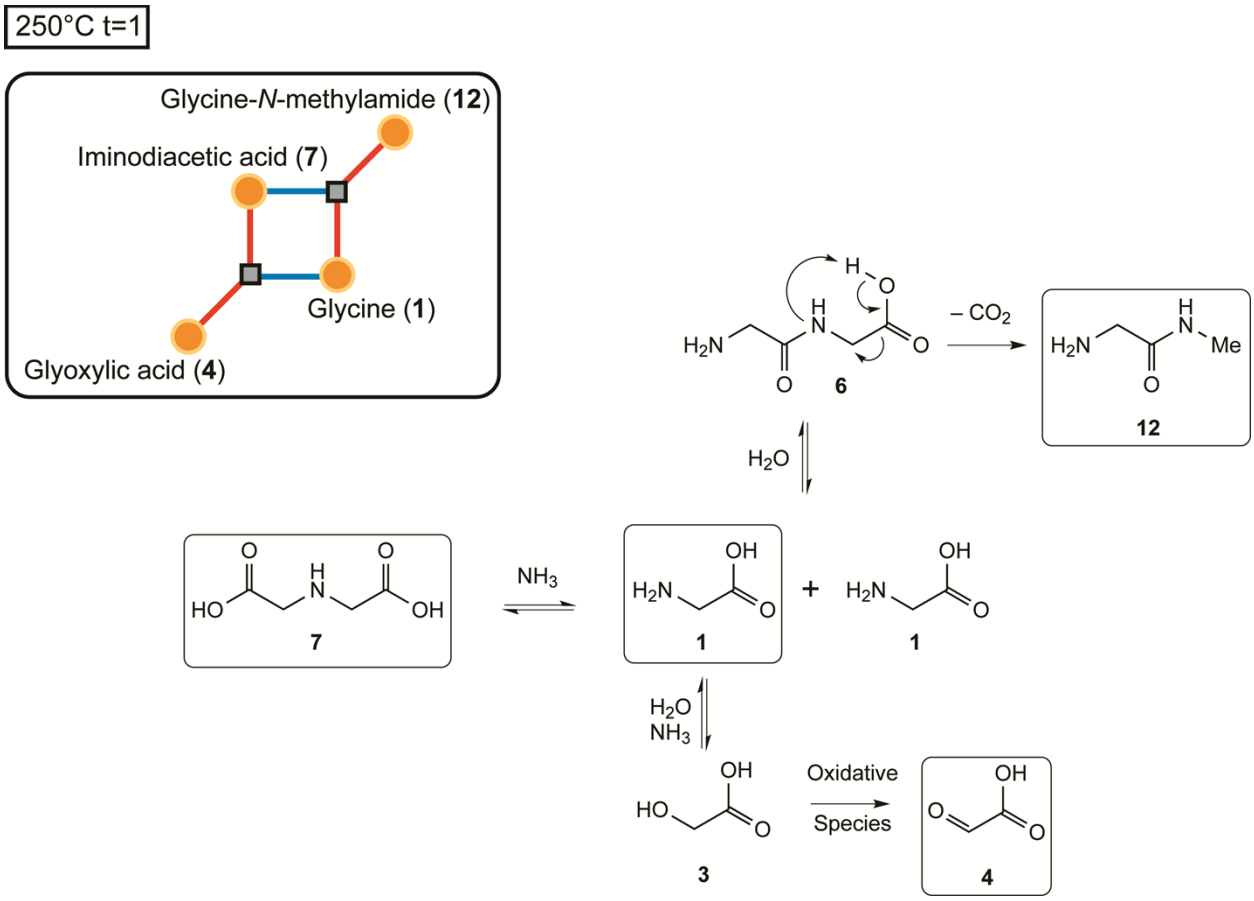
Supplementary Scheme 1. Interpreted reaction network of t_4 at 180 °C. The addition of NH_3 to iminodiacetic acid (**7**) induces its dissociation into two glycines (**1**). The nucleophilic substitution of glycine (**1**) by methylamine accompanied by the loss of NH_3 gives rise to sarcosine (**8**). The equilibrium between the previously formed (180 °C t=2, Scheme 1) *N*-carboxyglycine (**9**) and glycine (**1**) is displaced in favor of the condensation of two glycine species into 3,6-dihydropyrazine-2,5-diol (**17**) which can be additionally oxidized to obtain the thermodynamic aromatic compound (**18)** (pyrazine-2,5-diol). Methylamine can also be added to glycine (**1**) followed by the elimination of H_2O to furnish glycine-*N*-methylamide (**12**). To summarize, time point 4 at 180 °C is characterized by the consumption of iminodiacetic acid (**7**) and *N*-carboxyglycine (**9**) in order to accumulate sarcosine (**8**), glycine-*N*-methylamide (**12**) and pyrazine-2,5-diol (**18**).



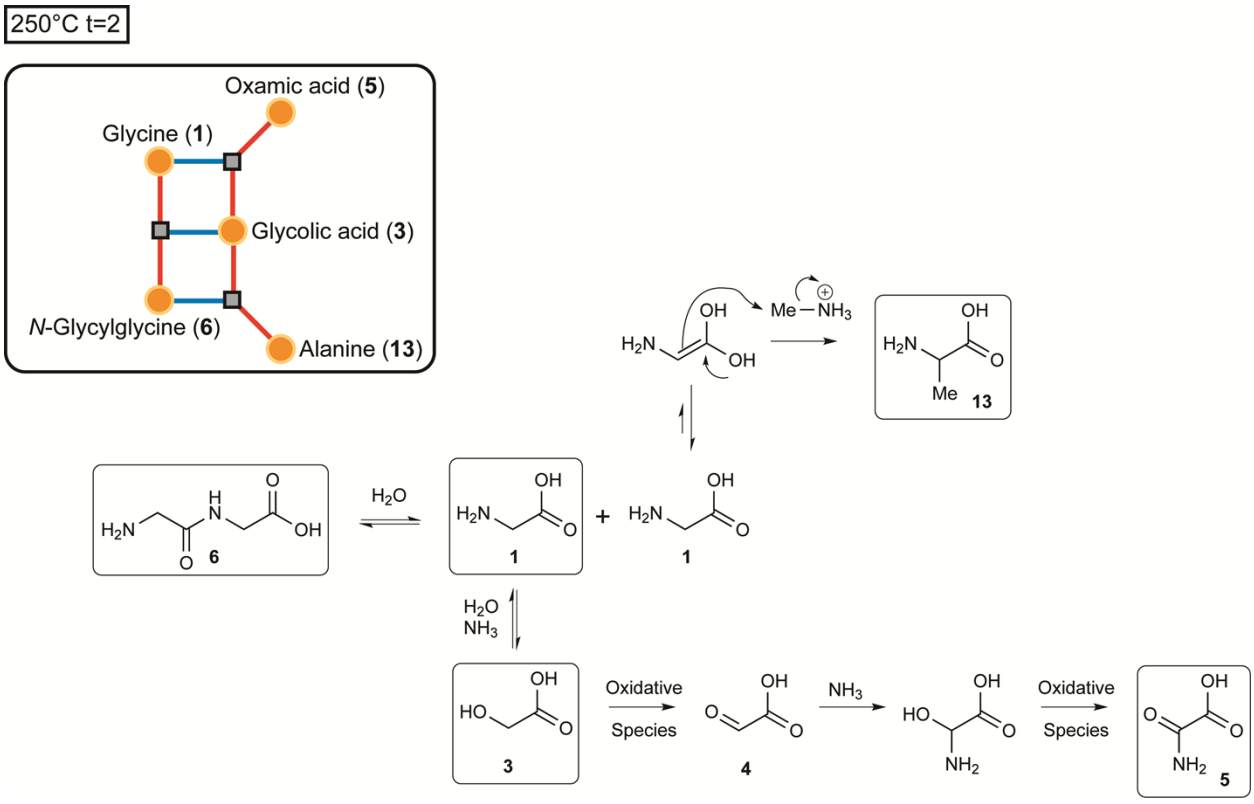
Supplementary Scheme 2. Interpreted reaction network of t_5 at 180 °C. The nucleophilic substitution of the starting material, glycine (**1**), by H_2O accompanied with the loss of NH_3 gives rise to glycolic acid (**3**). This carboxylic acid can then be oxidized by another compound present in the complex mixture into glyoxylic acid (**4**).



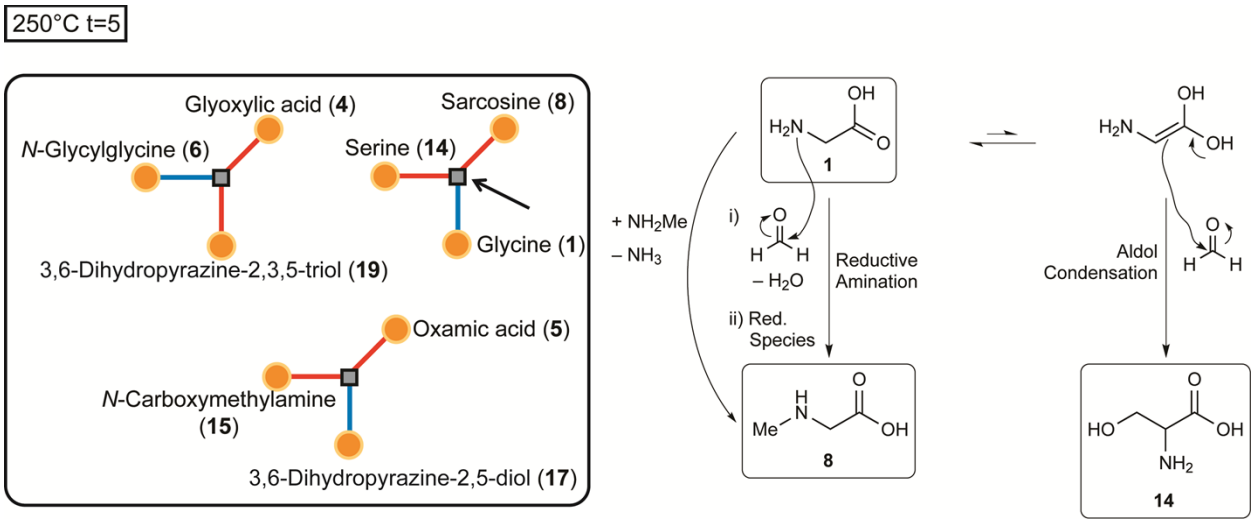
Supplementary Scheme 3. Interpreted reaction network of t_9 at 180 °C. The decarboxylation of glycine (**1**) leads to the formation of methylamine and carbon dioxide dissolved into carbonic acid (**2**). The two formed partners can then further react together. The addition of methylamine to carbonic acid (**2**), concurrent with the elimination of H_2O, produces *N*-carboxymethylamine (**15**).



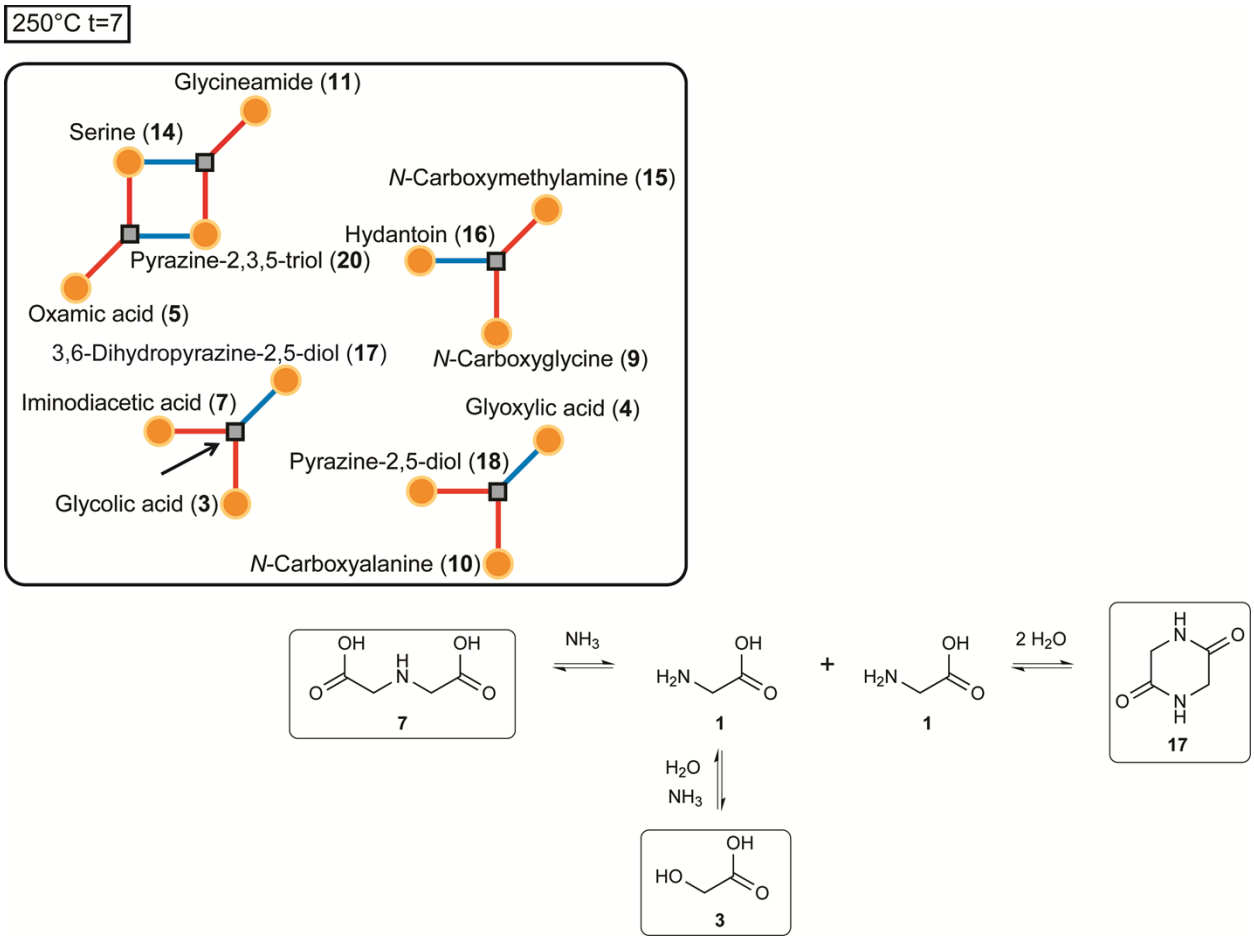
Supplementary Scheme 4. Interpreted reaction network of t_1 at 250 °C. An equilibrium between iminodiacetic acid (**7**) and two glycines (**1**) is observed and can displaced in favor of the formation of glycine-*N*-methylamide (**12**) and Glyoxylic acid (**4**). Glycine-*N*-methylamide (**12**) is obtained by condensation of glycines (**1**) into *N*-glycylglycine (**6**) followed by its decarboxylation. Concurrently, one of the two glycines (**1**), in equilibrium with iminodiacetic acid (**7**), is transformed into glycolic acid (**3**) by substitution with H_2O and loss of NH_3. Finally, glycolic acid (**3**) is further oxidized into glyoxylic acid (**4**) by an oxidative species present in the complex mixture. The succession glycine (**1**) to glycolic acid (**3**) to glyoxylic acid (**4**) was already observed at 180 °C, but at a much later time point (180 °C t=5, Supplementary Scheme 2). Altogether, this network can be interpreted as equilibrium between glycine (**1**) and iminodiacetic acid (**7**) and an accumulation of glyoxylic acid (**4**) and glycine-*N*-methylamide (**12**).



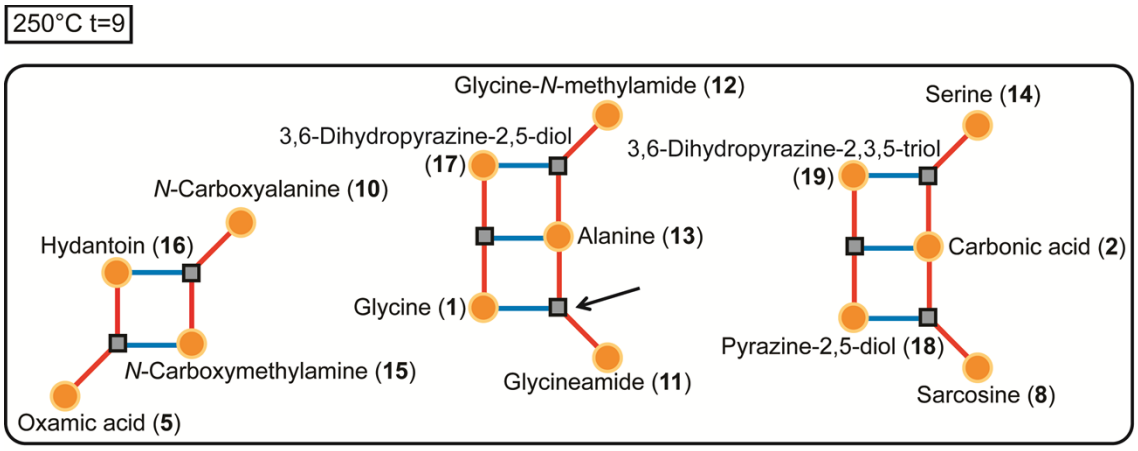
Supplementary Scheme 5. Interpreted reaction network of t_2 at 250 °C. The following sequence is proposed: *N*-Glycylglycine (**6**) and two glycines (**1**) are in equilibrium. Addition of H_2O to one of the glycines (**1**) concurrent with a loss of NH_3 gives rise to glycolic acid (**3**), further oxidized into glyoxylic acid (**4**). In order to attain oxamic acid (**5**), NH_3 is added to the aldehyde moiety of the glyoxylic acid and immediately oxidized before the consequent corresponding imine formation. In the meantime, the enol-tautomer of the glycine (**1**) obtained from the dissociation of *N*-glycylglycine (**6**) performs a nucleophilic substitution to a methylamine specie and after the loss of NH_3, the natural aminoacid alanine (**13**) is obtained. To conclude, the whole network can be understood as a equilibrium between three species, *N*-glycylglycine (**6**), glycine (**1**) and glycolic acid (**3**), and an accumulation of the two other partners, oxamic acid (**5**) and alanine (**13**).

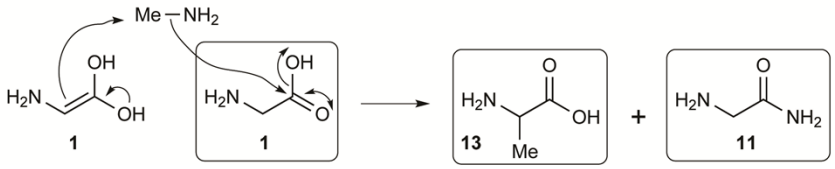


Supplementary Scheme 6. Interpreted reaction network of t_5 at 250 °C. Regarding the interpretation of the network between the starting material (**1**), serine (**14**) and sarcosine (**8**), we assume that two forms of glycine (**1**) under keto-enol equilibrium are present in the mixture. A fourth partner is involved as well, namely formaldehyde. Each of the two tautomers reacts with the same fourth partner but in a different manner to give two different products. The keto-tautomer of glycine (**1**) performs a reductive amination with formaldehyde and, with the help of a reductive specie present in the solution, sarcosine (**8**) is obtained. In the meanwhile, the enol-tautomer achieves an aldol condensation type reaction with formaldehyde so as to provide another natural aminoacid serine (**14**). Addition of methylamine to glycine (**1**) followed by a loss of NH_3 is a competitive way of obtaining sarcosine (**8**).

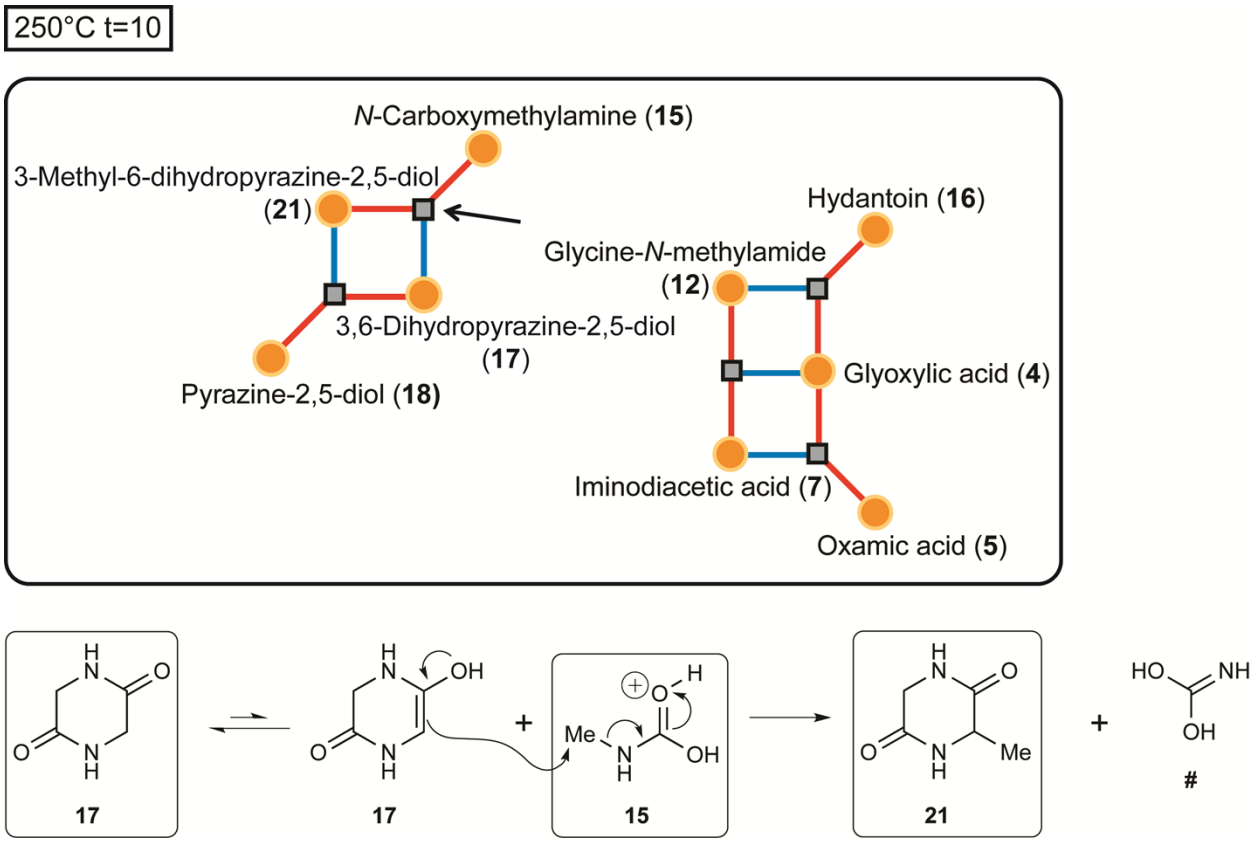


Supplementary Scheme 7. Interpreted reaction network of t_7 at 250 °C. As already observed in a previous time point, the equilibrium between iminodiacetic acid (**7**) and two glycines (**1**) (Supplementary Scheme 4) can be displaced in favor of the production of glycolic acid (**3**) and 3,6-dihydropyrazine-2,5-diol (**17**). The two glycines (**1**) can react again together and form the 3,6-dihydropyrazine-2,5-diol (**17**) accompanied by the elimination of two molecules of H_2O; or can be transformed into glycolic acid (**3**) by nucleophilic substitution of H_2O and loss of NH_3.

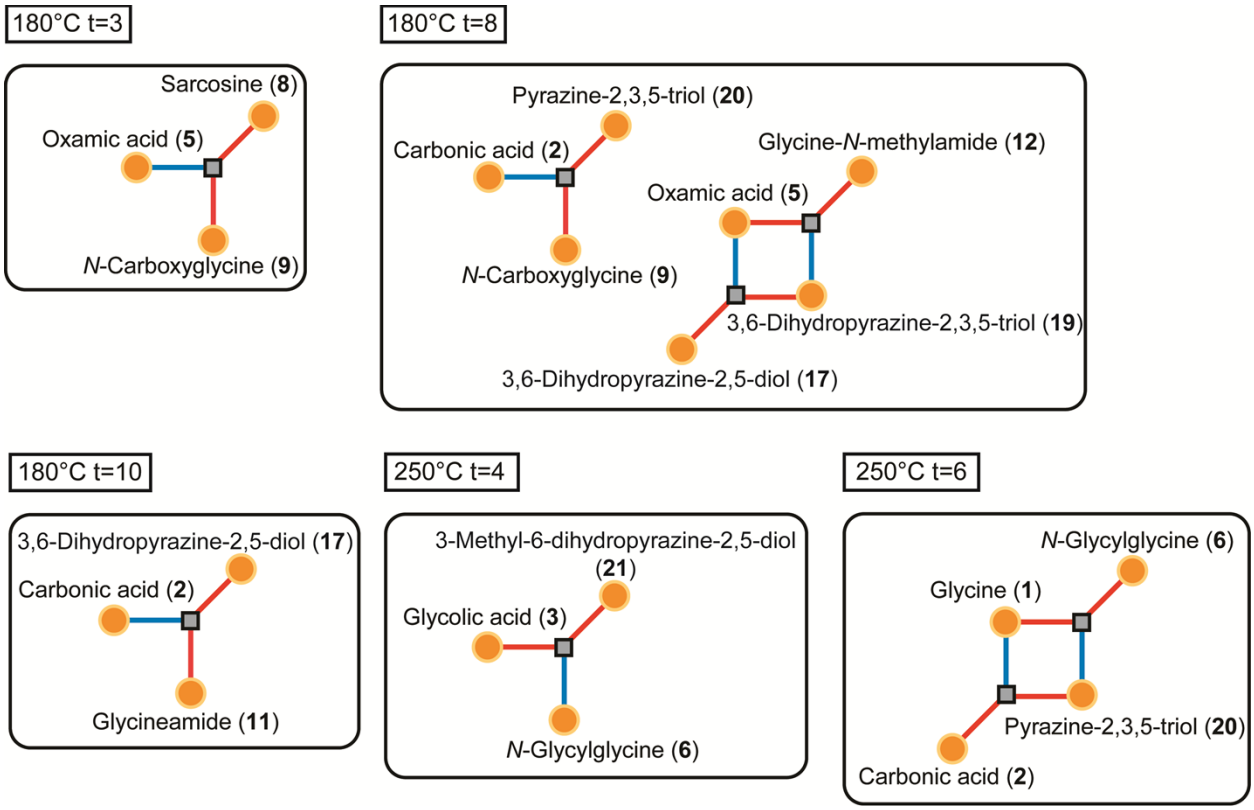




Supplementary Scheme 8. Interpreted reaction network of t_9 at 250 °C. Alanine (**13**) and glycineamide (**11**) are obtained after a reaction between two glycine species (**1**) and methylamine. The enol-tautomer of one glycine (**1**) substitutes the methylamine to produce alanine (**13**). The liberated NH_3 then performs an addition to another glycine (**1**) in order to give rise to glycineamide (**11**) after elimination of H_2O.



Supplementary Scheme 9. Interpreted reaction network of t_{10} at 250 °C. The enol-tautomer of the 3,6-dihydropyrazine-2,5-diol (**17**) executes a nucleophilic substitution to the *N*-carboxymethylamine (**15**) in order to obtain carbamic acid (#) and the 3-methyl-6-dihydropyrazine-2,5-diol (**21**).



Supplementary Figure 4. Compilation of additional non-interpreted reaction networks.

Compound ID	Compound	molecular formula	ID of matching NIST08 entry	ID of matching GMD entry	Unique Mass [mass fragment after electron impact ionization]	Retention Index	molecular formula of analyte	Number of methoxyamine moieties	Number of trimethylsilyl moieties	Confirmation by stable isotope labelling	Number of labelled carbon atoms (1,2-^{13}C-glycine)	Number of labelled nitrogen atoms (^{15}N-glycine)	ion formula	ion type	measured m/z [M+H]	error [mDa]	error [ppm]	mSigma	rank [hit list]
(1)	**Glycine**	C2H5NO2	NIST#: 78872	A114001	102	1110.76	C8H21NO2Si2		2		-	-	C8H22NO2Si2	[M+H]+	220.1182	0.1	0.7	412.0	1
										13C	2	-	C6H22NO2Si2^13C2	[M+H]+	222.1250	0.1	0.4	110.7	1
										15N	-	1	C8H22O2Si2^15N	[M+H]+	221.1160	-0.6	-2.6	70.4	1
	Glycine	C2H5NO2	NIST#: 28669	A133001	174	1302.68	C11H29NO2Si3		3	-	-	-	C11H30NO2Si3	[M+H]+	292.1578	0.1	0.3	76.8	1
										13C	2	-	C9H30NO2Si3^13C2	[M+H]+	294.1645	0.1	0.2	19.3	1
										15N	-	1	C11H30O2Si3^15N	[M+H]+	293.1546	0.3	1.1	18.2	1
(2)	**Carbonic acid**	CH2O3		A114002	133	1131.50	C10H26O3Si3		3	-	-	-	nd						
										13C	nd	nd	nd						
										15N	nd	nd	nd						
(3)	**Glycolic acid**	C2H4O3	NIST#: 352446	A106002	147	1062.89	C8H20O3Si2*		2	-	-	-	C11H29O3Si3	[M+H]+	293.1426	-0.7	-2.3	9.6	1
										13C	2	-	C9H29O3Si3^13C2	[M+H]+	295.1486	0.1	0.2	9.0	1
										15N	-	-	non labelled						
(4)	**Glyoxylic acid**	C2H2O3	NIST#: 24276	A100001	160	977.39	C6H13NO3Si	1	1	-	-	-	C6H14NO3Si	[M+H]+	176.0745	-0.7	-4.2	75.2	2
										13C	2	-	C4H14NO3Si^13C2	[M+H]+	178.0805	-0.1	-0.5	5.4	1
										15N	-	-	non labelled						
(5)	**Oxamic acid**	C2H3NO3		A125007	116	1255.25	C8H19NO3Si2*		2	-	-	-	C11H28NO3Si3	[M+H]+	306.1374	-0.3	-0.9	140.6	6
										13C	2	-	C9H28NO3Si3^13C2	[M+H]+	308.1440	-0.2	-0.5	4.0	1
										15N	-	1	C11H28O3Si3^15N	[M+H]+	307.1341	0.1	0.2	11.1	1
(6)	***N*-Glycylglycine**	C4H8N2O3	NIST#: 158696	A186012	174	1818.21	C13H32N2O3Si3		3		-	-	C13H33N2O3Si3	[M+H]+	349.1796	-0.3	-0.8	8.4	1
										13C	4	-	C9H33N2O3Si3^13C4	[M+H]+	353.1929	-0.1	-0.4	42.3	1
										15N	-	2	C13H33O3Si3^15N2	[M+H]+	351.1733	0.1	0.3	30.2	1
	***N*-Glycylglycine**	C4H8N2O3	NIST#: 158697	A186011	174	1800.39	C16H40N2O3Si4		4		-	-	C16H41N2O3Si4	[M+H]+	421.2191	-0.2	-0.6	12.9	2
										13C	4	-	C12H41N2O3Si4^13C4	[M+H]+	425.2322	0.1	0.3	5.8	1
										15N	-	2	C16H41O3Si4^15N2	[M+H]+	423.2128	0.1	0.3	18.7	1
(7)	**Iminodiacetic acid**	C4H7NO4		A151020	232	1521.37	C13H31NO4Si3		3	-	-	-	C13H32NO4Si3	[M+H]+	350.1634	-0.2	-0.5	2.5	1
										13C	4	-	C9H32NO4Si3^13C4	[M+H]+	354.1771	-0.4	-1.0	68.0	1
										15N	-	1	C13H32O4Si3^15N	[M+H]+	351.1603	0.1	0.3	8.0	1
(8)	**Sarcosine**	C3H7NO2	NIST#: 79236	A113004	116	1124.95	C9H23NO2Si2		2	-	-	-	C9H24NO2Si2	[M+H]+	234.1339	0.1	0.6	4.5	1
										13C	3	-	C6H24NO2Si2^13C3	[M+H]+	237.1444	-0.3	-1.4	11.5	1
										15N	-	1	C9H24O2Si2^15N	[M+H]+	235.1311	-0.1	-0.4	10.3	1
(9)	***N*-Carboxyglycine**	C3H5NO4		A147006	218	1462.19	C12H29NO4Si3		3	-	-	-	C12H30NO4Si3	[M+H]+	336.1479	-0.2	-0.5	5.2	1
										13C	2	-	C10H30NO4Si3^13C2	[M+H]+	338.1539	0.5	1.4	72.9	1
										15N	-	1	C12H30O4Si3^15N	[M+H]+	337.1445	0.2	0.6	6.8	1
(10)	*N*-Carboxyalanine	C4H7NO4		A142001	116	1395.20	C10H23NO4Si2		2	-	-	-	nd						
										13C	nd	nd	nd						
										15N	nd	nd	nd						
(11)	**Glycineamide**	C2H6N2O		A145025	174	1455.80	C11H30N2OSi3*		3	-	-	-	C14H39N2OSi4	[M+H]+	363.2137	-0.3	-0.8	187.7	15
										13C	2	-	C12H39N2OSi4^13C2	[M+H]+	365.2202	-0.1	-0.2	19.6	1
										15N	-	2	C14H39OSi4^15N2	[M+H]+	365.2069	0.6	1.5	26.1	1
(12)	Glycine-*N*-methylamide	C3H8N2O		A140020	174	1409.35	C9H24N2OSi2		2		-	-	C9H25N2OSi2	[M+H]+	233.1498	0.2	0.9	2.5	1
										13C	3	-	C6H25N2OSi2^13C3	[M+H]+	236.1600	0.1	0.4	14.3	1
										15N	-	2	C9H25OSi2^15N2	[M+H]+	235.1441	-0.1	-0.2	51.2	1
	Glycine-*N*-methylamide	C3H8N2O		A144018	174	1448.15	C12H32N2OSi3		3		-	-	C12H33N2OSi3	[M+H]+	305.1903	-0.8	-2.5	11.7	1
										13C	3	-	C9H33N2OSi3^13C3	[M+H]+	308.1991	0.5	1.6	5.3	1
										15N	-	2	C12H33OSi3^15N2	[M+H]+	307.1832	0.4	1.3	24.6	1
(13)	**Alanine**	C3H7NO2	NIST#: 79162	A110001	116	1087.41	C9H23NO2Si2		2	-	-	-	C9H24NO2Si2	[M+H]+	234.1343	-0.2	-1.0	18.4	1
										13C	3	-	C6H24NO2Si2^13C3	[M+H]+	237.1446	-0.5	2.1	1.1	1
										15N	-	1	C9H24O2Si2^15N	[M+H]+	235.1316	-0.5	-2.3	6.3	1
(14)	**Serine**	C3H7NO3	NIST#: 333284	A128001	116	1254.21	C9H23NO3Si2		2	-	-	-	C9H24NO3Si2	[M+H]+	250.1286	0.3	1.2	113.1	4
										13C	nd	nd	interference						
										15N	nd	nd	interference						
	Serine	C3H7NO3	NIST#: 333286	A138001	204	1352.85	C12H31NO3Si3		3	-	-	-	C12H32NO3Si3	[M+H]+	322.1684	0.0	0.0	4.8	1
										13C	3	-	C9H32NO3Si3^13C3	[M+H]+	325.1784	0.1	0.4	6.1	1
										15N	-	1	C12H32O3Si3^15N	[M+H]+	323.1654	0.0	0.1	16.4	1
(15)	*N*-Carboxymethylamine	C2H5NO2		A109003	204	1094.69	C8H21NO2Si2		2	-	-	-	C8H22NO2Si2	[M+H]+	220.1182	0.1	0.6	7.9	1
										13C	2	-	C6H22NO2Si2^13C2	[M+H]+	222.1250	0.1	0.4	10.5	1
										15N	-	1	C8H22O2Si2^15N	[M+H]+	221.1160	-0.6	-2.6	9.0	1
(16)	Hydantoine	C3H4N2O2		A141014	229	1415.12	C9H20N2O2Si2		2		-	-	C9H21N2O2Si2	[M+H]+	245.1134	0.2	0.7	11.4	2
										13C	3	-	C6H21N2O2Si2^13C3	[M+H]+	248.1239	-0.2	-0.9	7.6	1
										15N	-	2	C9H21O2Si2^15N2	[M+H]+	247.1076	0.0	0.2	10.5	1
(17)	**3,6-Dihydropyrazine-2,5-diol**	C4H6N2O2	NIST#: 151999	A142014	116	1424.08	C10H22N2O2Si2		2		-	-	C10H23N2O2Si2	[M+H]+	259.1300	-0.7	-2.7	22.1	1
										13C	4	-	C6H23N2O2Si2^13C4	[M+H]+	263.1426	0.1	0.4	4.4	1
										15N	-	2	C10H23O2Si2^15N2	[M+H]+	261.1229	0.5	1.8	6.7	1
(18)	**Pyrazine-2,5-diol**	C4H4N2O2	NIST#: 151998	A135011	241	1355.99	C10H20N2O2Si2		2		-	-	C10H21N2O2Si2	[M+H]+	257.1144	-0.8	-3.1	16.0	1
										13C	4	-	C6H21N2O2Si2^13C4	[M+H]+	261.1272	-0.2	-0.6	15.9	1
										15N	-	2	C10H21O2Si2^15N2	[M+H]+	259.1078	-0.1	-0.4	65.5	1
(19)	**3,6-Dihydropyrazine-2,3,5-triol**	C4H6N2O3	NIST#: 156301	A159011	116	1596.64	C13H30N2O3Si3		3		-	-	C13H31N2O3Si3	[M+H]+	347.1638	-0.1	-0.4	28.4	3
										13C	4	-	C9H31N2O3Si3^13C4	[M+H]+	351.1774	-0.3	-0.8	14.8	1
										15N	-	2	C13H31O3Si3^15N2	[M+H]+	349.1578	-0.3	-0.9	11.7	1
(20)	**Pyrazine-2,3,5-triol**	C4H4N2O3	NIST#: 156300	A156013	241	1560.76	C13H28N2O3Si3		3		-	-	C13H29N2O3Si3	[M+H]+	345.1483	-0.2	-0.7	15.2	2
										13C	4	-	C9H29N2O3Si3^13C4	[M+H]+	349.1620	-0.5	-1.4	8.5	1
										15N	-	2	C13H29O3Si3^15N2	[M+H]+	347.1422	-0.1	-0.3	13.2	1
(21)	**3-Methyl-3,6-dihydropyrazine-2,5-diol**	C5H8N2O2	NIST#: 333700	A137012	257	1379.49	C11H24N2O2Si2		2		-	-	C11H25N2O2Si2	[M+H]+	273.1451	-0.2	-0.8	14.9	1
										13C	5	-	C6H25N2O2Si2^13C5	[M+H]+	278.1618	-0.2	-0.6	4.7	1
										15N	-	2	C11H25O2Si2^15N2	[M+H]+	275.1389	0.1	0.3	7.1	1
(22)	Methylpyrazine-2,5-diol	C5H6N2O2		-	nd	1357.38	C11H22N2O2Si2		2		-	-	C11H23N2O2Si2	[M+H]+	271.1291	0.1	0.5	13.8	1
										13C	nd	nd	interference		276.1467				
										15N	nd	nd	interference		273.1254				
(23)	-	C4H8N2O		-	nd	1289.09	C7H16N2OSi		1		-	-	C7H17N2OSi	[M+H]+	173.1108	-0.3	-1.9	3.6	1
										13C	4	-	C3H17N2OSi^13C4	[M+H]+	177.1239	0.0	0.1	7.3	1
										15N	-	2	C7H17OSi^15N2	[M+H]+	175.1048	-0.2	-1.4	9.8	1
(24)	-	C4H6N2O		-	nd	1680.42	C7H14N2OSi		1		-	-	C7H15N2OSi	[M+H]+	171.0948	0.0	-0.1	4.3	1
										13C	4	-	C3H15N2OSi^13C4	[M+H]+	175.1083	-0.1	-0.5	0.5	1
										15N	-	2	C7H15OSi^15N2	[M+H]+	173.0890	-0.1	-0.7	6.7	1

underline (confirmed by authenticated reference substance)

bold (confirmed by mass spectral match to the GMD or NIST08 libraries)

*APCI mass and spectrum indicates an additional trimethylsilyl moiety.

nd (not determined)

Supplementary Table 1. Characterization of glycine HTR products in complex mixture without purification by GC-MS profiling approaches. Analyses were performed by GC-MS hyphenated to electron impact ionization and time of flight MS (GC-EI-TOF-MS) and by paired analyses of the same samples using GC-MS hyphenated to atmospheric-pressure chemical ionization (APCI) and high mass accuracy time of flight MS (APCI-TOF-MS). The supplementary table contains compound numbering (cf. Fig. 1), name, and molecular formula. The access identifiers of the EI-TOF-MS fragmentation spectra uploaded into the Golm Metabolome Database (http://gmd.mpimp-golm.mpg.de/search.aspx) and respective matching entries of the NIST08 mass spectral library (http://webbook.nist.gov/chemistry/) are listed with the unique, *i.e.* the selective mass fragments used for relative quantification, and with retention indices based on *n*-alkane standardization of a 5%-phenyl-95%-dimethylpolysiloxane capillary column. The right part of the table summarizes the elucidation and confirmation of molecular formula by APCI-TOF-MS of HTR products from non-labeled glycine, 1,2-$^{13}C_2$-glycine, and ^{15}N-glycine. The expected molecular formula of the trimethylsilylated and in one case methoxyaminated analyte are compared to the molecular formula deduced from exact mass and isotopomer distribution measurements of the APCI-TOF-MS technology. This table contains 2 additional glycine HTR products. These products were not further investigated as co-eluting compounds and lower sensitivity of the EI-TOF-MS compared to the APCI-TOF-MS technology did not allow the isolation of a pure and paired EI-TOF-MS fragmentation spectrum.

Supplementary Table 2. The chemical profiles of glycine HTR products obtained at 100 bar pressure and either 0–7.16 min at 180 °C or 0–3.23 min at 250 °C. The table provides arbitrary GC-EI-TOF-MS abundances of unique mass fragments of each of the component chemicals. Unique masses are listed in Table S1. Note that the unique masses and respective

reaction kinetic data can be used for descriptive statistics and the reversed engineering of reaction networks without knowledge of compound identity and structure.

Additional Data file S1 (separate file). Full set of GC-EI-TOF-MS mass fragmentation spectra of all non-labeled glycine HTR products and respective isotopomer spectra of the products obtained under identical conditions from ^{15}N-, 1-^{13}C-, 2-^{13}C- and 1,2-$^{13}C_2$-labeled glycine. This data file is formatted for upload into the AMDIS or NIST mass spectra search and comparison softwares (http://chemdata.nist.gov/mass-spc/ms-search/). Copies of each mass spectrum can be retrieved from the Golm Metabolome Database (http://gmd.mpimp-golm.mpg.de/).

References

1 Fitzpatrick, J. H. J. & Hopgood, D. Metal-ion catalyzed decarboxylation - kinetics and mechanism of oxidative decarboxylation of copper(II) complexes of aminomalonic acid in aqueous-solution. *Inorg. Chem.* **13**, 568–574 (1974).

2 Farthing, A. C. Synthetic polypeptides. 1. Synthesis of oxazolid-2-5-diones and a new reaction of glycine. *J. Chem. Soc.*, 3213–3217 (1950).

3 Kopka, J. Current challenges and developments in GC-MS based metabolite profiling technology. *J. Biotechnol.* **124**, 312–322 (2006).

4 Matsuda, F. *et al.* MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. *Plant J.* **57**, 555–577 (2009).

5 Lisec, J., Schauer, N., Kopka, J., Willmitzer, L. & Fernie, A. R. Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat. Protoc.* **1**, 387–396 (2006).

6 Luedemann, A., Strassburg, K., Erban, A. & Kopka, J. TagFinder for the quantitative analysis of gas chromatography - mass spectrometry (GC-MS)-based metabolite profiling experiments. *Bioinformatics* **24**, 732–737 (2008).

7 Cuadros-Inostroza, A. *et al.* TargetSearch - a Bioconductor package for the efficient preprocessing of GC-MS metabolite profiling data. *BMC Bioinformatics* **10**, 428 (2009).

8 Kopka, J. *et al.* GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* **21**, 1635–1638 (2005).

9 Hummel, J., Strehmel, N., Selbig, J., Walther, D. & Kopka, J. Decision tree supported substructure prediction of metabolites from GC-MS profiles. *Metabolomics* **6**, 322–333 (2010).

10 Strehmel, N., Hummel, J., Erban, A., Strassburg, K. & Kopka, J. Retention index thresholds for compound matching in GC-MS metabolite profiling. *J. Chromat. B, Analyt. Technol. Biomed. Life Sci.* **871**, 182–190 (2008).

11 Bergström, K., Gurtler, J. & Blomstrand, R. Trimethylsilylation of amino acids. 1. Study of glycine and lysine TMS Derivatives with gas-liquid chromatography-mass spectrometry. *Anal. Biochem.* **34**, 74–87 (1970).

12 Leimer, K. R., Rice, R. H. & Gehrke, C. W. Complete mass-spectra of per-trimethylsilylated amino-acids. *J. Chromatogr.* **141**, 355–375 (1977).

13 Hempel, S., Koseska, A., Nikoloski, Z. & Kurths, J. Unraveling gene regulatory networks from time-resolved gene expression data - a measures comparison study. *BMC Bioinformatics* **12**, 292 (2011).

14 Sweetlove, L. J., Fell, D. & Fernie, A. R. Getting to grips with the plant metabolic network. *Biochem. J.* **409**, 27–41 (2008).

15 Klamt, S., Haus, U. U. & Theis, F. Hypergraphs and cellular networks. *PLoS Comput. Biol.* **5**, e1000385 (2009).

16 Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B Methodol.* **58**, 267–288 (1996).

17 Smyth, G. in *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (eds R. Gentleman *et al.*) 397–420 (Springer, 2005).

18 Hennig, C., fpc: Flexible procedures for clustering. R package version 2.0-3. http://CRAN.R-project.org/package=fpc (2010).

19 Calinski, T. & Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.* **3**, 1–27 (1974).

20 Stacklies, W., Redestig, H., Scholz, M., Walther, D. & Selbig, J. pcaMethods - a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **23**, 1164–1167 (2007).

21 Klingler, D., Berg, J. & Vogel, H. Hydrothermal reactions of alanine and glycine in sub- and supercritical water. *J. Supercrit. Fluids* **43**, 112–119 (2007).