Published in final edited form as:

Angew Chem Int Ed Engl. 2015 August 17; 54(34): 10030–10034. doi:10.1002/anie.201502868.

Chemical Evolution of a Bacterial Proteome

Dr. Michael Georg Hoesl^[a], M. Sc. Stefan Oehm^[a], Dr. Patrick Durkin^[a], Dr. Elise Darmon^[f], Dr. Lauri Peil^[c], Hans-Rudolf Aerni, PhD^[d], Prof. Dr. Juri Rappsilber^{[b],[c]}, Prof. Jesse Rinehart, PhD^[d], Prof. Dr. David Leach^[f], Prof. Dr. Dieter Söll^[e], and Prof. Nediljko Budisa^[a]

Nediljko Budisa: nediljko.budisa@tu-berlin.de

^[a]Institut für Chemie, Technische Universität Berlin Müller-Breslau-Str. 10, 10623 Berlin

^[b]Institut für Biotechnologie, Technische Universität Berlin Gustav-Meyer-Allee 25, 13355 Berlin

^[c]Wellcome Trust Centre for Cell Biology, School of Biological Sciences, University of Edinburgh 4.17 Michael Swann Building, Edinburgh EH9 3BF

^[d]Systems Biology Institute Yale University, West Haven, CT 06516, USA

^[e]Department of Molecular Biophysics and Biochemistry Yale University, New Haven, CT 06520, USA Department of Chemistry, Yale University, New Haven, CT 06520, USA

^[f]Institute of Cell Biology, School of Biological Sciences, University of Edinburgh, Kings Buildings, Edinburgh EH9 3BF

Abstract

We have changed the amino acid set of the genetic code of *Escherichia coli* by evolving cultures capable of growing on the synthetic non-canonical amino acid L- β -(thieno[3,2-b]pyrrolyl)-alanine ([3,2]Tpa) as a sole surrogate for the canonical amino acid L-tryptophan (Trp). A long-term cultivation experiment in defined synthetic media resulted in the evolution of cells capable of surviving Trp $\rightarrow$ [3,2] Tpa substitutions in their proteomes in response to the 20,899 TGG codons of the *E. coli* W3110 genome. These evolved bacteria with new-to-nature amino acid composition are capable of robust growth in the complete absence of Trp. Our experimental results illustrate an approach for the evolution of synthetic cells with alternative biochemical building blocks.

Graphical Abstract

Towards synthetic bacteria: An evolution experiment in synthetic medium led to quantitative Tryptophan to *L*- β -(thieno[3,2-*b*]pyrrolyl)-alanine substitution in response to 20,899 TGG codons in bacterium *Escherichia coli* W3110. The evolved bacteria were capable of robust growth in the complete absence of Trp without significant adverse effects on the cellular survival.

Keywords

Tryptophan analog; *Escherichia coli*; Continuous evolution; Synthetic biology; Genetic code translation

Translational ambiguity as an underlying mechanism behind genetic code flexibility is a “sweet spot” for the development and evolution of alternative genetic codes.^[1–3] Based on these findings we ventured on the next step towards the experimental diversification of the genetic code: the complete substitution of a canonical amino acid (cAA) by a non-canonical one (ncAA). In contrast to stop codon suppression focused approaches,^[4–6] we set out to incorporate ncAAs in a proteome-wide manner. Although previous work in this direction has been conducted,^[7–9] the absence of sophisticated analytics and proteomics tools hindered conclusive evidence of full proteome-wide replacement with a non-canonical analog.

Trp is believed to be evolution’s latest addition to the genetic code^[10] and demands the highest metabolic synthesis cost of all proteinogenic amino acids. Therefore, it has relatively low abundance in proteins (~1% i.e. about 20,000 residues in the whole *E. coli* proteome^[11]) and is encoded by a single codon (UGG). It possesses special biophysical properties which allow for its participation in numerous interactions ($\pi \rightarrow \pi$ stacking, hydrogen bonding, cation- π interactions). Therefore, Trp plays a major role in protein stability and folding, and participates in mediation processes such as receptor-ligand interactions or enzyme-substrate binding. Thus, substitution of Trp with other cAAs can often result in misfolded proteins and inactive enzymes, ultimately causing cell death. However, Trp’s diverse and rich indole chemistry offers numerous possibilities for potential analogs which might take over Trp’s function after cellular and genomic bacterial adaptation. In other words, Trp substitution in proteins with related aromatic systems seems to be plausible without causing totally detrimental effects on the structural and functional integrity of the cell. In the 1950s, the possibility for incorporation of 7-azatryptophan and 2-azatryptophan into proteins was reported.^[12,13] However, until recently Trp residues have been mainly replaced in single recombinant proteins by various non-canonical aza-, fluoro-, amino-, and hydroxyl-tryptophan analogs.^[14]

On the proteomic scale, Trp-auxotrophic *Bacillus subtilis* strains selected for growth on 4-fluorotryptophan (4FTrp) as a sole substitute for Trp have been reported.^[8] However, they were grown in ‘rich’ synthetic medium supplemented with vitamins, nucleobases and amino acids. These additives can metabolically compensate for potentially deleterious effects of 4FTrp on protein function and lessen the selection pressure for a completely functional modified proteome. Furthermore, these strains were generated by the use of a mutagenic agent hampering conclusive genotype analyses. A different strategy to adapt *E. coli* to 4FTrp consisted of long-term continuous batch culturing by serial dilution^[15] in pure mineral

medium.^[9] During continuous cultivation, the initial Trp content of the medium was decreased stepwise while the concentration of 4FTrp was constantly raised. At the end of the long-term cultivation experiment, the resulting strain was able to grow, but only very poorly (end OD_{600nm} ≈ 0.1) on 4FTrp. However, the impurity of the commercial 4FTrp preparation (i.e. residual traces of Trp) was responsible for a significant residual presence of Trp (~90 nM) in the final culture medium. In the present study, we set ourselves the goal to select an *E. coli* strain capable of growing with a Trp analog in pure mineral medium without any traces of canonical Trp.

To achieve this purpose, we chose to use [3,2]Tpa as the Trp analog (Figure 1) in our evolution experiment since its synthetic route excludes any indole or Trp intermediate (see SI). [3,2]Tpa is known to be translationally active and incorporation into single target proteins has already been demonstrated.^[16] In this earlier study we had already observed that the replacement of the benzene ring of Trp with thiophene was not only well tolerated by the auxotrophic cells, but [3,2]Tpa was also used to some extent as a substrate for cellular growth.^[16]

In the present study we unambiguously confirmed by *in vitro* and *in vivo* activation and aminoacylation studies that [3,2]Tpa is a good substrate for *E. coli* tryptophanyl-tRNA synthetase (Section 3 in the SI).

Suitable auxotrophic strains are a prerequisite to apply selection pressure for amino acid analog usage by the cell. We have chosen the *E. coli* K12 W3110 derivative CGSC# 7679.^[17] In this strain, the whole Trp biosynthesis pathway is removed (*trpLEDCBA*). In Trp-auxotrophic *E. coli* strains, Trp and its analogs enter the cell via transporter-mediated uptake.^[18] This mechanism might be a ready-made target for the cell to shut down analog uptake and avoid disadvantageous consequences of incorporation into proteins. Therefore, our experimental setup was based on indole (Ind) and β-thieno[3,2-*b*]pyrrole ([3,2]Tp) (Figure 1), which enter bacterial cells by passive diffusion through the membrane.^[19] To convert these precursors intracellularly into Trp and [3,2]Tpa respectively, we equipped CGSC# 7679 with the plasmid pSTB7^[20], which harbors the *Salmonella typhimurium* Trp synthase (TrpBA), an enzyme known to efficiently convert [3,2]Tp into [3,2]Tpa. Our initial strain configuration *E. coli* K12 W3110 (*trpLEDCBA*) <pSTB7> was designated MT0.

We used New Minimal Medium (NMM) throughout the evolution experiment.^[21] In the beginning, the medium contained a set of all canonical amino acids except Trp (NMM19) to remove the pathways of amino acid biosynthesis from the initial evolutionary landscape encountered by the cells. Additionally, supplying amino acids leads to higher culture densities, facilitating the reduction of indole concentration, and to shorter generation times that are doubtlessly advantageous for initial adaptation. The amino acids were gradually removed at later stages of the experiment to end up with a complete mineral medium and glucose as the sole carbon source (NMM0).

MT0 cells did not grow in the presence of [3,2]Tp (25 μM) without added indole. However, additional supply of only 1 μM of indole was sufficient to obtain a final OD_{600nm} of 0.8–0.9. Starting from this configuration, we conducted a long term evolution experiment by serially

re-inoculating our *E. coli* culture in shaking flasks. The cells were gradually forced to lose their Trp dependence by continuously reducing the amount of indole from the initial 1 μM to zero while constantly keeping 25 μM of [3,2]Tp in the medium (Figure 2).

Interestingly, already after 106 serial passages (164 days of cultivation) the population was able to grow in NMM19 without any further indole supplementation. After around 150 passages, maximum optical culture densities stopped fluctuating and stabilized around 1.3 OD_{600nm}. With passage 170 (289 cultivation days) we started a stepwise withdrawal of amino acids. To identify amino acids not essential for culture growth on [3,2]Tp, we grouped the 19 canonical amino acids into their metabolic blocks and removed them in a combinatorial manner (Figure 2). Subsequently, within these blocks, some individual amino acids were removed from the medium one by one (Section 5 in the SI). A total of six steps of amino acid removal were required. It has to be pointed out that we were only able to remove the two last amino acids, methionine and histidine, from the medium when we decreased the culture temperature from standard 37 °C to 30 °C. We hypothesise that [3,2]Tpa incorporation led to misfolding of enzymes important in these amino acid biosynthesis pathways.

Finally, after 264 passages we obtained an *E. coli* population capable of growing in mineral glucose medium (NMM0) supplemented with [3,2]Tp. The full adaptation to growth on [3,2]Tp took place in the time frame of 506 days. At that stage, the culture grew to 1.6 OD_{600nm} in around two days (Figure 2). The purity and genetic identity of the evolution culture was continuously controlled on agarose plates and by PCR amplification of key genomic regions within the deleted Trp operon and plasmid pSTB7 (Sections 6 and 7 in the SI). The culture was free of contamination and maintained auxotrophy for Trp/[3,2]Tpa throughout the whole evolution experiment.

At different time points of the experiment, we have isolated robustly growing clones from the cultures to document the different stages of evolution in our continuous culture (MT strains). The endpoint isolate termed MT20 is capable of growth at 30 °C on NMM0 containing only [3,2]Tp while keeping its preference for indole as well. Comparing the growth kinetics of MT0 and MT1, an isolate originated from the middle stage of the experiment, the difference in generation times in LB at 37°C increases slightly from 24.4 to 28.8 minutes. It should also be mentioned that the overnight OD_{600nm} decreased from 3.5 in MT0 to 2.2 in MT1. These results indicate that the MT1 strain partially accumulated mutations lowering its relative fitness in LB full medium in comparison to the original MT0 strain.

We also performed initial cell morphological investigations and have found that cells which evolved to grow in the presence of [3,2]Tp present a different phenotype, as shown in Figure 3. In general, evolved MT1 cells exhibit increased length (~1 μm) and reduced width (~0.1 μm). Curiously, MT1 cells grown in NMM19 medium were thinner when grown in presence of indole. This may be a consequence of amino acid starvation processes triggered by the absence of the canonical Trp.^[22] We are currently performing comprehensive microscopic and genome analyses in order to shed light on possible physiological mechanisms underlying these changes in fitness.

To quantify [3,2]Tpa incorporation at Trp positions in the proteome and to exclude the possibility of trace Trp presence or genetic strain reversion, we carried out highly sensitive analytics using gas chromatography and MS/MS measurements. We chose re-grown MT isolates in their respective culture conditions as samples for these analyses (Table 1 and sections 5 and 9 in the SI).

MS/MS analyses were conducted using individually grown cultures in two different laboratories using distinct MS/MS techniques (one example is shown in Figure S5). Both analyses successfully provided direct evidence for a full Trp $\rightarrow$ [3,2]Tpa substitution throughout the whole proteome. We analyzed in detail the 20 most abundant peptides containing a single Trp position and emerging from different proteins (Section 8 in the SI). Computer aided MS/MS data analyses showed that cell cultures grown exclusively on [3,2]Tp incorporated the ncAA into their proteome with no Trp identification in peptides. In addition, heterologous expression of eGFP in [3,2]Tp containing medium resulted in completely labeled protein as verified by ESI-MS analysis (Section 7 in the SI).

The application of very sensitive GC/MS analyses enabled us to unambiguously verify the complete liberation of the proteome from Trp (Table 1). Interestingly, in cultures of isolates MT1 and MT12, traces of Trp were still detected in the proteome even though indole was no longer added to the culture media. Trace amounts of Trp may be explained by a contamination of medium preparations. This should not be surprising as the majority of commercially available amino acids are preparations from biological fermentative processes.

With the full removal of all amino acids from the medium, the residual Trp traces completely disappeared (MT16 – MT20). Therefore, in the frame of the detection limits of our methodology and the available instrumentation, these GC/MS analyses demonstrate the complete proteome-wide absence of Trp traces in the evolved cells. It should be kept in mind that we could not provide direct evidence for the presence of [3,2]Tpa via GC/MS analysis because it is readily degraded in the hydrolysis step of the amino acid analysis. This is not surprising as thieno-surrogates of Trp are less stable (e.g., the aromatic delocalization energy of benzene is 36 kcal/mol whereas thiophene and thiazole have only 29 kcal/mol and 25 kcal/mol, respectively)^[23].

We have generated an organism whose proteomic composition changed from initial Trp to [3,2]Tpa during a long term evolution experiment without significant negative impact on its cellular survival. This uncovered a high evolutionary plasticity that has allowed the substitution of all the tryptophan in the proteome of *E. coli*. We are currently analyzing genomes of evolved strains and anticipate (based on previous work with fluorinated Trp analogs^[8,9]) that only a relatively small number of proteins should be detrimentally affected by [3,2]Tpa incorporation.^[24] Genetic encoding in living systems is based on highly standardized chemistry composed of the same number of “letters” or nucleotides as informational polymers (DNA, RNA) and the twenty α -amino acids as basic building blocks for proteins. The universality of the genetic code enables the horizontal transfer of genes across biological taxa which affords a high degree of standardization and interconnectivity. Thus, all deep chemistry changes within living systems tend to be generally lethal.^[25–27]

In this context, one of the great challenges for 21st century bioscience is the development of a strategy for expanding the standard basic chemical repertoire of living cells to achieve biocontainment by man-made or naturally evolved changes in the genetic code.^[28–30] Recently, the non/canonical chemical barrier was recently crossed by evolving an *E. coli* strain with a genome in which thymine was replaced 5-chlorouracil.^[31] This is a first important example in the construction of biological systems composed of xeno-nucleic acids.

Based on the notion that between 30 and 40 sense codons are adequate to encode the genetic information of an organism,^[32] a large number of sense codons (>20) may be available for recoding with ncAAs. Current ‘genetic code expansion’ approaches^[33], programmed to reassign UAG or UGA stop codons, are geared to produce modified proteins containing one or multiple ncAAs with fluorescent or chemically reactive groups for a host of *in vivo* or *in vitro* applications. However, proteome-wide replacements of amino acids are not practical by this route. Our long-term *E. coli* evolution experiment resulted in an organism in which all 20,899 UGG codons were ‘recoded’ by substituting the supply of the original Trp-tRNA^{Trp} with a ncAA-tRNA (i.e. [3,2]Tpa-tRNA^{Trp}) formed by the same aminoacyl-tRNA synthetase. This example of a proteome-wide amino acid replacement represents a significant ‘genetic code expansion’ and a real step towards a synthetic organism.

Taken together, approaches to alter the meaning of the genetic information stored in DNA as informational polymer by changing the chemistry of the polymer (i.e., xeno-nucleic acids) or by changes in the genetic code have already yielded successful examples.^[34] In the future this should enable the partial or full redirection of the biological information flow to generate ‘new’ version(s) of the genetic code derived from the ‘old’^[25] biological world.

Experimental Section

The experimental section can be found in the SI.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Laure Prat for advice on acid urea gels and northern blot analysis. Traudl Wenger gave excellent assistance in strain passaging, phenotype and genotype analysis. We thank Nina Bach (TU München) for ESI-MS analysis of eGFP-H6[[3,2]Tpa]. Special thanks should be given to Julia Forstner for her logistic support. We acknowledge financial support from the Einstein–Foundation supported ARTCODE-consortium, FP7 EU-funded METACODE-Consortium, UniCat Excellence Cluster of TU Berlin, from DFG Forschergruppe 1805 (Ribosome Dynamics in Regulation of Speed and Accuracy of Translation), the graduate school “Fluorine as a key element” (GRK1582) of DFG, the National Institute for General Medical Sciences (to DS) and the Defense Advanced Research Projects Agency (contracts N66001-12-C-4020 and N66001-12-C-4211 to DS).

References

1. Jukes TH, Osawa S. Comp Biochem Physiol B. 1993; 106:489–94. [PubMed: 8281749]
2. Osawa S, Jukes TH, Watanabe K, Muto A. Microbiol Rev. 1992; 56:229–64. [PubMed: 1579111]
3. Moura GR, Paredes Ja, Santos MaS. FEBS Lett. 2010; 584:334–41. [PubMed: 19941859]

4. Noren CJ, Anthony-Cahill SJ, Griffith MC, Schultz PG. *Science*. 1989; 244:182–188. [PubMed: 2649980]
5. Furter R. *Protein Sci*. 1998; 7:419–426. [PubMed: 9521119]
6. Dumas A, Lercher L, Spicer CD, Davis BG. *Chem Sci*. 2015; 6:50–69.
7. Cowie DB, Cohen GN. *Biochim Biophys Acta*. 1957; 26:252–61. [PubMed: 13499359]
8. Wong JTF. *Proc Natl Acad Sci U S A*. 1983; 80:6303–6306. [PubMed: 6413975]
9. Bacher JM, Ellington AD. *J Bacteriol*. 2001; 183:5414–25. [PubMed: 11514527]
10. Fournier GP, Alm EJ. *J Mol Evol*. 201510.1007/s00239-015-9672-1
11. Lajoie MJ, Rovner AJ, Goodman DB, Aerni HR, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA, et al. *Science*. 2013; 342:357–60. [PubMed: 24136966]
12. Pardee AB, Prestidge LS, Shore VG. *Biochim Biophys Acta*. 1956; 21:406–7. [PubMed: 13363942]
13. Brawerman G, Ycas M. *Arch Biochem Biophys*. 1957; 68:112–177. [PubMed: 13435899]
14. Budisa N, Pal PP. *Biol Chem*. 2004; 385:893–904. [PubMed: 15551863]
15. Blount ZD, Borland CZ, Lenski RE. *Proc Natl Acad Sci U S A*. 2008; 105:7899–906. [PubMed: 18524956]
16. Budisa N, Alefelder S, Bae JH, Golbik R, Minks C, Huber R, Moroder L. *Protein Sci*. 2001; 10:1281–92. [PubMed: 11420430]
17. Yanofsky C, Horn V. *J Bacteriol*. 1995; 177:4451–6. [PubMed: 7543478]
18. Yanofsky C, Horn V, Gollnick P. *J Bacteriol*. 1991; 173:6009–17. [PubMed: 1917834]
19. Piñero-Fernandez S, Chimere C, Keyser UF, Summers DK. *J Bacteriol*. 2011; 193:1793–8. [PubMed: 21296966]
20. Kawasaki H, Bauerle R, Zon G, Ahmed SA, Miles EW. *J Biol Chem*. 1987; 262:10678–83. [PubMed: 3112150]
21. Budisa N, Steipe B, Demange P, Eckerskorn C, Kellermann J, Huber R. *Eur J Biochem*. 1995; 230:788–796. [PubMed: 7607253]
22. Grossman N, Ron EZ, Woldringh CL. 1982; 152:35–41.
23. Bird CW. *Tetrahedron*. 1985; 41:1409–1414.
24. Yu ACS, Yim AKY, Mat WK, Tong AHY, Lok S, Xue H, Tsui SKW, Wong JTF, Chan TF. *Genome Biol Evol*. 2014; 6:629–41. [PubMed: 24572018]
25. Marliere P. *Syst Synth Biol*. 2009; 3:77–84. [PubMed: 19816802]
26. Schmidt M. *Bioessays*. 2010; 32:322–31. [PubMed: 20217844]
27. Acevedo-Rocha CG, Budisa N. *Angew Chem Int Ed Engl*. 2011; 50:6960–2. [PubMed: 21710510]
28. Mandell DJ, Lajoie MJ, Mee MT, Takeuchi R, Kuznetsov G, Norville JE, Gregg CJ, Stoddard BL, Church GM. *Nature*. 201510.1038/nature14121
29. Rovner AJ, Haimovich AD, Katz SR, Li Z, Grome MW, Gassaway BM, Amiram M, Patel JR, Gallagher RR, Rinehart J, et al. *Nature*. 201510.1038/nature14095
30. Campbell JH, O'Donoghue P, Campbell AG, Schwientek P, Sczyrba A, Woyke T, Söll D, Podar M. *Proc Natl Acad Sci U S A*. 2013; 110:5540–5. [PubMed: 23509275]
31. Marlière P, Patrouix J, Döring V, Herdewijn P, Tricot S, Cruveiller S, Bouzon M, Mutzel R. *Angew Chem Int Ed Engl*. 2011; 50:7109–7114. [PubMed: 21710668]
32. O'Donoghue P, Ling J, Wang YS, Söll D. *Nat Chem Biol*. 2013; 9:594–8. [PubMed: 24045798]
33. Chin JW. *Annu Rev Biochem*. 2014; 83:379–408. [PubMed: 24555827]
34. Budisa N. *Curr Org Chem*. 2014; 18:936–943.

Figure 1.

Structures of 1: tryptophan (Trp), 2: indole (Ind), 3: β-thieno[3,2-*b*]-pyrrole ([3,2]Tp), and L-β-(thieno[3,2-*b*]pyrrolyl)alanine ([3,2]Tpa).

Figure 2. Adaptation of *E. coli* strain CGSC# 7679 <pSTB7> to [3,2]Tpa

The upper scale indicates the passage count while the lower scale shows the time in days. In each passage, OD_{600nm} was measured. The growth temperature was decreased from 37°C to 30°C while going from NMM2s to NMM1. The different medium compositions are indicated on the right side and are described in details in section 4 in the SI.

Figure 3.

Morphology and average distribution of cell length (A) and width (B) as measured by microscopy of MT0 and MT1. Strains were grown in NMM19 with either 100 μM of Indole (black dots $\rightarrow$ MT0; blue diamonds $\rightarrow$ MT1) or of [3,2]Tp (orange squares $\rightarrow$ MT1). Error bars represent the standard errors of the means of four experiments (with 350 cells studied by experiment). Inserts: examples of brightfield microscopy pictures of MT0 (0 μM [3,2]Tp / 100 μM Indole) and MT1 (0 μM Indole / 100 μM [3,2]Tp) cells grown in NMM19 and prepared on a 1% agarose-coated slide to be visualised by microscopy. The calibration bar indicates 5 μm . Details can be found in sections 9 and 10 in the SI.

Table 1

Proteome-wide [3,2]Tpa incorporation followed by mass spectrometric and amino acid analyses. The MT-strains were isolated at different time points throughout the long term evolution experiment (Figure 2) and represent increasing stages of adaptation. Please note that [3,2]Tpa is the amino acid version of the Trp analog, which is produced introcellularly by the strain. In contrast, [3,2]Tp is the Ind analog version which is supplied to the medium. An extended version of this table with additional intermediates can be found in the SI (Table S5).

Strain ^[a]	MT0	MT0	MT1	MT20
MS/MS^[b]				
Trp [%]	99.5 ± 1.2	99.7 ± 0.5	0.6 ± 0.8	0.5 ± 0.8
[3,2]Tpa [%]	0.5 ± 1.2	0.6 ± 0.5	99.4 ± 0.8	99.5 ± 0.8
Amino acid analysis^[c]				
Trp [%]	100 ± 3.8	98.7 ± 3.2	1.6 ± 0.1	^[d]
Medium				
Aa concent ^[e]	19	19	19	0
Ind [μM]	100	100	0	0
[3,2]Tp [μM]	0	25	25	100

^[a] Strain details are provided in section 4 and 6 in the SI;

^[b] Measured peptide occurrence of [3,2]Tpa and Trp measured by MS/MS. 20 of the most abundant tryptophan containing peptides were chosen for quantification;

^[c] GC/MS analysis of hydrolysed proteome samples derivatized by esterification and amidation.

^[d] no peak observed in the chromatogram;

^[e] Amino acid content of the medium. A detailed overview of the medium composition can be found in section 4 in the SI.