Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Sci.*, 2023, **14**, 11718

All publication charges for this article have been paid for by the Royal Society of Chemistry

Received 31st May 2023
Accepted 21st September 2023

DOI: 10.1039/d3sc02773b

rsc.li/chemical-science

A cellular platform for production of C_4 monomers†

Matthew A. Davis,‡[a] Vivian Yaci Yu,‡[a] Beverly Fu,[b] Miao Wen,[b] Edward J. Koleski,[b] Joshua Silverman,[c] Charles A. Berdan,[b] Daniel K. Nomura[abd] and Michelle C. Y. Chang*[abe]

Living organisms carry out a wide range of remarkable functions, including the synthesis of thousands of simple and complex chemical structures for cellular growth and maintenance. The manipulation of this reaction network has allowed for the genetic engineering of cells for targeted chemical synthesis, but it remains challenging to alter the program underlying their fundamental chemical behavior. By taking advantage of the unique ability of living systems to use evolution to find solutions to complex problems, we have achieved yields of up to ~95% for three C_4 commodity chemicals, *n*-butanol, 1,3-butanediol, and 4-hydroxy-2-butanone. Genomic sequencing of the evolved strains identified *pcnB* and *rpoBC* as two gene loci that are able to alter carbon flow by remodeling the transcriptional landscape of the cell, highlighting the potential of synthetic pathways as a tool to identify metabolic control points.

Introduction

Living systems rely on a dynamic and complex network of chemical reactions to carry out the tasks needed to coordinate cellular growth and maintenance, allowing transformation of simple carbon sources into the thousands of molecules needed for life. As such, cells possess an enormous synthetic potential that can be engineered for targeted chemical synthesis, enabling the reduction of traditionally multi-step synthetic routes into a single fermentation step that can be carried out in water and under ambient temperature and pressure. Their ability to utilize building blocks including sugars from renewable plant biomass, CO_2, and CH_4 for biosynthesis opens a route to shifting industrial chemical production away from its traditional reliance on petrochemical feedstocks towards a universal fermentation platform.

A major challenge in the development of cell-based chemical synthesis is that the reaction network used to produce target compounds is also used to carry out basic cell functions. These reactions are thus subject to many levels of regulation in order to maintain the necessary coordination between parts of the metabolic network.[1–3] In particular, key hubs of the metabolic map, such as the central carbon pathways of glycolysis and the tricarboxylic acid (TCA) cycle, form many connections with the rest of the network and are difficult to manipulate as their behavior is affected by multiple inputs and outputs.[4] As a result, the construction of high-yielding pathways can be difficult to achieve as evolution drives the cell to direct carbon flux to cell growth and biomass in competition with engineered biosynthesis.

Since these central carbon pathways are closely tied to cell state, they are correspondingly subject to homeostatic mechanisms to ensure robustness to change. Therefore, many simultaneous alterations are needed to rationally engineer carbon flow to insufficiently active nodes.[5–8] However, an advantage that living systems provide is that evolution can be used to solve this multi-dimensional problem if product titers can be tied to cell growth.[9,10] In this work, we demonstrate the design and evolution of synthetic pathways to selectively produce three industrially-relevant C_4 compounds: 1,3-butanediol (butylene glycol, BDO), 4-hydroxy-2-butanone (HB), and *n*-butanol (Fig. 1A). These three compounds are used for various purposes, ranging from pharmaceutical precursors (BDO and HB) to a drop-in gasoline replacement (*n*-butanol).[11–13] In particular, BDO can be used as a humectant or solvent for a variety of different high-value products, as well as a co-monomer for production of various polymers. These three compounds can also be further dehydrated to produce the C_4 monomers 1,3-butadiene (from BDO),[14] methyl vinyl ketone (from HB),[15] and 1-butene (from *n*-butanol).[16] Using a genetic selection, the yields of these pathways were improved from 11–20% to near

[a]*Department of Molecular & Cellular Biology, University of California, Berkeley, CA 94720-3200, USA. E-mail: mcchang@berkeley.edu*
[b]*Department of Chemistry, University of California, Berkeley, CA 94720-1460, USA*
[c]*Calysta, 1900 Alameda de las Pulgas Suite 200, San Mateo, CA 94404, USA*
[d]*Department of Nutritional Sciences & Toxicology, University of California, Berkeley, CA 94720-3104, USA*
[e]*Department of Chemical & Biomolecular Engineering, University of California, Berkeley, CA 94720-1462, USA*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3sc02773b

‡ Matthew A. Davis and Vivian Yaci contributed equally to this work.

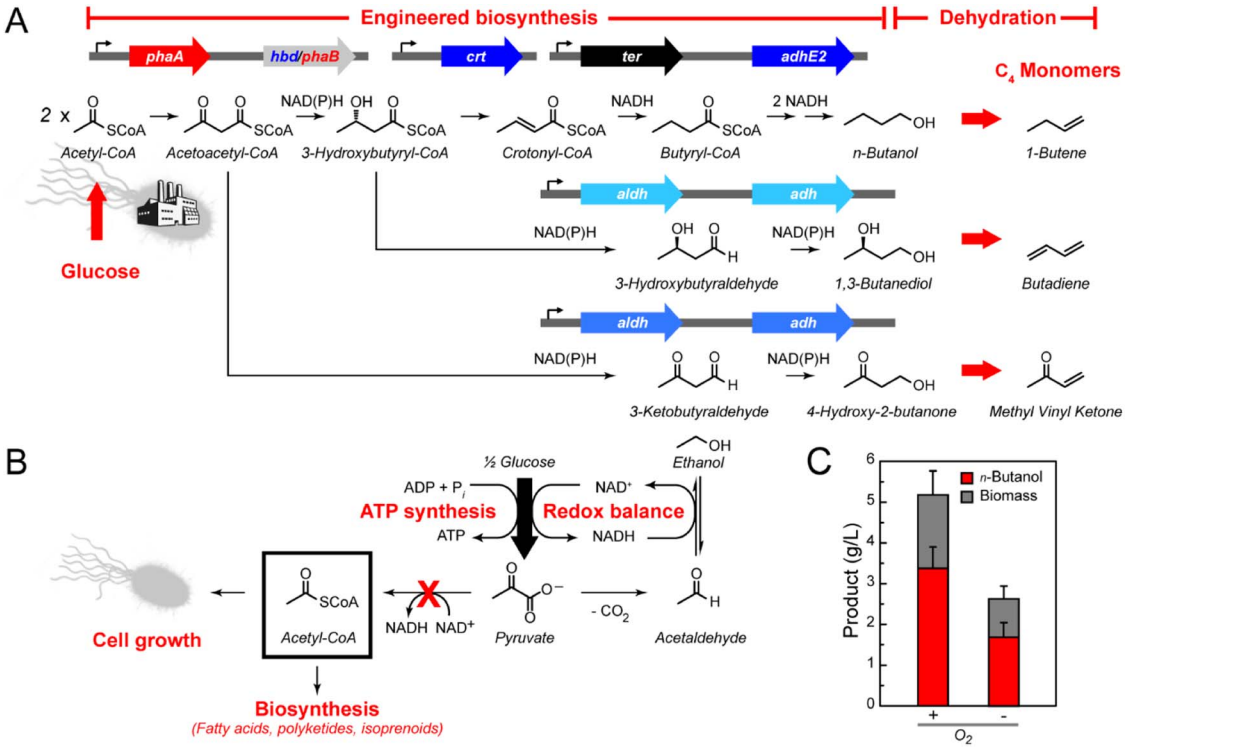

Fig. 1 Synthetic pathways for production of C_4 monomers. (A) Design of a platform for production of C_4 monomers based on *n*-butanol formation. Identification of selective aldehyde and alcohol dehydrogenases enables the formation of three different C_4 products from glucose: *n*-butanol, 1,3-butanediol, and 4-hydroxy-2-butanone *via* engineered microbes. Chemical dehydration of these compounds produces the industrially-relevant C_4 monomers 1-butene, butadiene, and methyl vinyl ketone, respectively. *phaA*, acetoacetyl-CoA synthase; *phaB*, R-specific NAD(P)H-dependent acetoacetyl-CoA dehydrogenase; *hbd*, S-specific NADH-dependent acetoacetyl-CoA dehydrogenase; *crt*, crotonase; *ter*, *trans*-enoyl-CoA reductase; *adhE2*, bifunctional aldehyde/alcohol dehydrogenase; *aldh*, aldehyde dehydrogenase; *adh*, alcohol dehydrogenase. Genes derived from the poly(hydroxy)alkanote pathway of *Ralstonia eutrophus* are labeled in red. Genes derived from the acetone–butanol–ethanol pathway of *Clostridium acetobutylicum* are labeled in royal blue. Gene from *Treponema denticola* is labeled in black. (Light blue *aldh* and *adh* genes denote their general function.) (B) Anaerobic fermentation pathways can operate at near quantitative yields in the absence of O_2. Under these conditions, substrate-level phosphorylation pathways such as glycolysis serve as the only route to ATP synthesis but require the use of NAD^+. In Baker's yeast (*Saccharomyces cerevisiae*), decarboxylation of pyruvate and subsequent reduction to ethanol allow for the stoichiometric regeneration of NAD^+ and is required for cell survival. Because of the low ATP yield under anaerobic growth, cell growth as well as flux to anabolic pathways utilizing the key building block, acetyl-CoA, are greatly reduced. As a result, acetyl-CoA is not readily available for the downstream biosynthesis of a broad range of target compounds during anaerobic growth. (C) Production of *n*-butanol and biomass in *E. coli* DH1 containing a synthetic butanol pathway (pBT33-Bu1 pCWori-ter.adhE2 pBBR1-aceEF.lpd) under aerobic and anaerobic conditions. Data are mean ± s.d. of biological replicates ($n = 3$).

quantitative yields. Genome sequencing of the evolved strains showed that two gene loci, *pcnB* and *rpoBC*, were mutated in the most successful daughter cells. Subsequent characterization demonstrated that mutations at these two loci are sufficient to capture the majority of the evolved phenotype and likely operate by large-scale shifts in the transcriptome. Taken together, these results highlight the possibility of synthetic pathways to be used not only for scalable chemical production but also as a platform for discovery and study of cellular function.

Results and discussion

Implementation of a genetic selection for C_4 production

Anaerobic cell culture is often preferred for industrial fermentations because limitations in culture oxygenation on large-scale can be eliminated and theoretical product yields can be increased.[17] Under anaerobic conditions, carbon assimilation pathways like glycolysis serve as the primary route for cellular ATP synthesis since the lack of oxygen as a terminal electron acceptor makes aerobic respiration unavailable (Fig. 1B).[9,18] The relatively low ATP yield from substrate-level phosphorylation then results in a minimal loss of carbon to competing cell growth or biomass accumulation.[19] In addition to 2 mol of ATP, 2 mol of NADH is also generated from 1 mol of glucose through glycolysis. Fermentative pathways provide a mechanism to oxidize NADH to NAD^+, which is needed for glycolysis to remain operational. High flux through fermentation pathways is thus driven by cell survival. Lactate and ethanol production provide the paradigms for this process, resulting in rapid and near-quantitative yield from sugar *via* pyruvate (Fig. 1B).

Like ethanol and lactate, the C_4 alcohol, *n*-butanol, can serve to balance glucose fermentation because its biosynthesis recycles the four NADH produced per molecule of glucose. However, a major challenge for production of longer-chain target

compounds is that they typically require building blocks from pathways downstream of glycolysis and whose intracellular concentrations are regulated at many levels. One of the most important of these building blocks is acetyl coenzyme A (CoA), which is a two-carbon intermediate that serves as a central point of many metabolic decision points.[20,21] Acetyl-CoA synthesis and usage are tightly controlled with flux dropping under anaerobic conditions as both biosynthesis and cell growth are greatly reduced during fermentative growth (Fig. 1B). Indeed, *n*-butanol titers are greatly lowered when our first-generation *Escherichia coli* production strain was cultured anaerobically (Fig. 1C).[11] In order to reduce carbon flow to competing native pathways, the major fermentation pathways were knocked out of *E. coli* DH1 to generate DH1 *ΔldhA ΔadhE ΔfrdBC ΔpoxB ΔackA-pta* (DH1Δ5),[10,22] a selection strain that would require production of *n*-butanol for growth under anaerobic conditions (Fig. 2A, ESI Table S1 and Fig. S1†).[13] In order to provide a means to increase flux to acetyl-CoA, the pyruvate dehydrogenase complex (PDHc, *aceEF-lpd*) was overexpressed for the oxidative decarboxylation of pyruvate to produce acetyl-CoA and NADH to stoichiometrically balance *n*-butanol production from glucose.

The fermentation knockout strain DH1Δ5 was found competent to grow under anaerobic conditions when the synthetic *n*-butanol pathway consisting of *phaA*, *hbd*, *crt*, *ter*, and *adhE2* was expressed (Fig. 1). However, increased ethanol production, rather than *n*-butanol, was observed (Fig. S2†). The most likely source of ethanol was the promiscuity of AdhE2, which is a bifunctional aldehyde-alcohol dehydrogenase that produces both *n*-butanol and ethanol from the respective two-step reduction of butyryl-CoA and acetyl-CoA in its native host (Fig. 2A).[23,24] In this case, direct reduction of two equivalents of acetyl-CoA to ethanol would also regenerate the necessary four NAD^+ per glucose and creates a short circuit in the pathway to circumvent *n*-butanol production (Fig. 2A). Biochemical analysis of different AdhE2 constructs was carried out in order to

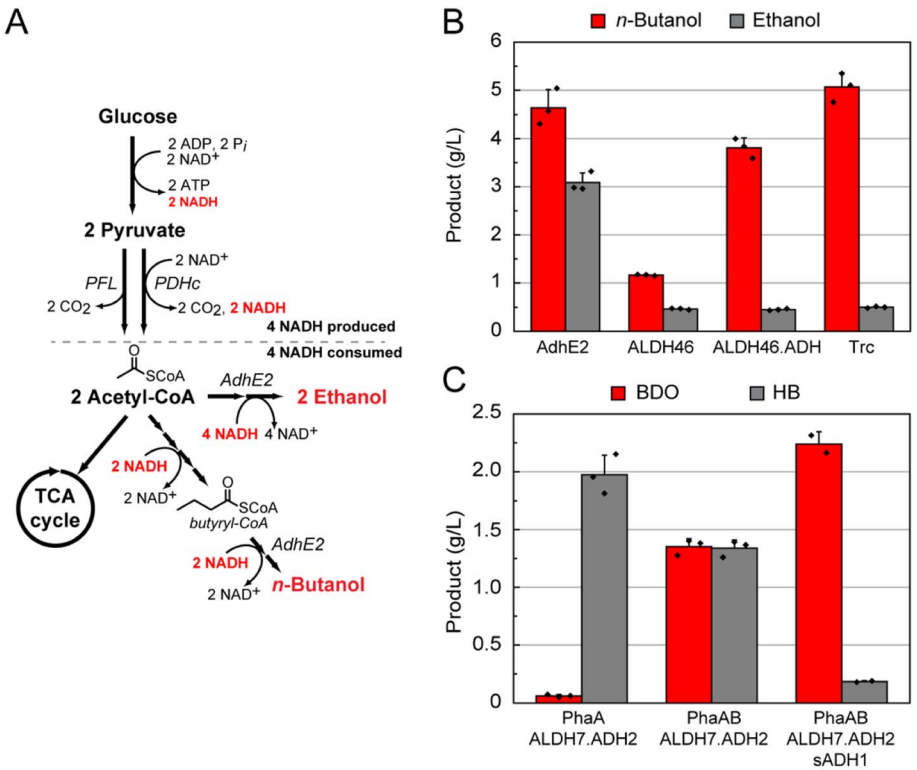

Fig. 2 Production of C_4 monomer precursors in engineered *E. coli*. (A) Design of a host for the anaerobic production of target compounds from acetyl-CoA. Deletion of the major fermentation pathways of *E. coli* in DH1Δ5 allows the synthetic *n*-butanol pathway to be the major mechanism for balanced NAD^+ regeneration *via* the acetyl-CoA intermediate. However, the promiscuity of AdhE2 towards acetyl-CoA and butyryl-CoA leads to ethanol fermentation as a pathway short-circuit that also maintains stoichiometric redox balance. (PDHc, pyruvate dehydrogenase complex; PFL, pyruvate formate lyase). (B) Screening of AdhE, ALDH, and ADH candidates in *E. coli* DH1Δ5 pBBR1-aceEF.lpd pBT33-Bu2 yields a C_4-selective fermentation pathway under anaerobic conditions. When AdhE2 is included, high levels of ethanol are produced along with the target *n*-butanol product. Replacement with ALDH46 reduces ethanol production to background levels but concomitantly drops *n*-butanol titers. Addition of the ADH domain from AdhE2 and tuning the promoter for expression allows for high n-butanol yields with very little ethanol being formed. All strains were grown anaerobically in TB with 2.5% (w/v) glucose media for 3 days post induction. (AdhE2, pCWori-ter-adhE2; ALDH46, pCWori-ter.aldh46; ALDH46.ADH, pCWori-ter-aldh46.ADHAdhE2; Trc, pCWori.trc-ter-aldh46.ADHAdhE2). (C) Screening of ALDH, ADH, and sADH candidates in *E. coli* pT533-phaA/phaAB pCWori.trc-ter-ALDH.ADH led to identification of the ALDH7.ADH2 pair for production of HB and BDO under anaerobic conditions. In the absence of PhaB, HB is selectively produced. Addition of PhaB leads to a 1 : 1 ratio of both products being formed. The inclusion of sADH then allows for HB to be converted to BDO. All strains were grown anaerobically in TB with 2.5% (w/v) glucose media for 5 days post induction. Data are mean ± s.d. of biological replicates ($n = 3$).

assess the selectivity of each domain for the C_4 and C_2 substrates (Fig. S3†). Although there is a ten-fold preference for butyryl-CoA over acetyl-CoA, the higher k_{cat}/K_M for the C_4 substrate arises directly from a 10-fold lower K_M (10 ± 1 μM) with no change in k_{cat} within error. Given that the K_M for acetyl-CoA (100 ± 10 μM) is well within the expected physiological range (0.5–1.0 mM), it is likely that AdhE2 is capable of producing both *n*-butanol and ethanol at competitive rates under intracellular conditions. Although the mechanism of substrate channeling between the aldehyde dehydrogenase (ALDH) and alcohol dehydrogenase (ADH) domains is not yet fully understand, the high K_Ms (4.0–4.5 mM) measured for the aldehyde intermediate imply that substrate selection is controlled by the ALDH domain. Thus, we set out to identify enzymes that could efficiently carry out the reduction of butyryl-CoA while excluding acetyl-CoA.

Identifying C_4-selective dehydrogenases

Enzymes that tailor acyl-CoA substrates are typically permissive to a broad range of chain lengths, making the exclusion of a smaller substrate, like acetyl-CoA, challenging.[25,26] We therefore initiated a search for acylating ALDH candidates with characterized substrate selectivity and found three C_2-specific bifunctional AdhE2 homologs, four C_4-specific monofunctional ALDHs, and one atypical C_2-specific ALDH (Fig. S4†). These sequences were arranged in a biased phylogenetic tree with branching guided by characterized substrate preference (Fig. S4†). Next, the entire ALDH gene family from the Pfam database was assembled into a second phylogenetic tree based on the first tree (Fig. S4†) to generate a full family tree of <1200 sequences.[27] This tree was dominated by monofunctional ALDH sequences (67%) as the majority of ALDH domains found in the Pfam database are derived from standalone enzymes. Approximately 40 mutations were selected from the natural sequence diversity in the C_4 branch and used to design 95 AdhE2 variants (Table S1†). Each variant contained 3–5 mutations, and every mutation was present in multiple variants, ensuring that each mutation can be evaluated in multiple contexts. These AdhE2 variants were synthesized, cloned into expression vectors, and co-transformed into DH1Δ5 with the appropriate butanol production plasmids for *in vivo* screening (Fig. S5†). Around two-thirds of the variants (66 variants, 69%) remained active; however, only mild improvements in the *n*-butanol : ethanol ratio were observed.

Given the modest gains using this approach, we turned our attention to screening wild-type ALDH sequences falling within the C_4-selective branch since it seemed likely that the sequence information derived mostly from monofunctional ALDHs did not accurately predict the selectivity of their bifunctional counterparts. The C_4 branch of the tree was widely sampled to incorporate the full diversity of this branch in a small number of sequences comprising 15 bifunctional AdhE2 homologs and 18 monofunctional ALDHs (Table S1†). We found that all bifunctional enzymes except one yielded lower *n*-butanol:ethanol ratios compared to AdhE2 (Fig. S5†), possibly because a large majority of sequenced AdhE2 homologs are thought to be involved in ethanol generation and likely display a natural preference for acetyl-CoA. In contrast, 15 out of 16 monofunctional ALDHs produced more *n*-butanol than ethanol, possibly because the natural substrate range is larger with this enzyme family (Fig. S6†). Out of these monofunctional ALDHs, ALDH46 was selected as the final candidate for butyryl-CoA reduction. We next sought to improve overall *n*-butanol production to the levels observed using AdhE2 (Fig. 2B). We reasoned that the bottleneck was the reduction of butyraldehyde to *n*-butanol due to the absence of the ADH domain. In our biochemical studies of AdhE2, we characterized a truncation mutant consisting of solely the ADH domain of AdhE2 (ADH_{AdhE2}) that surprisingly showed an order of magnitude decrease in K_M for butyraldehyde to 300 ± 50 μM. We thus supplemented our pathway containing ALDH46 with the ADH_{AdhE2} domain, which more than doubled *n*-butanol titers. Increasing the promoter strength for expression of ALDH46-ADH_{AdhE2} then improved *n*-butanol titers in the two-protein system beyond that observed in the original bifunctional AdhE2-dependent pathway, with no ethanol production above background (Fig. 2B).

Developing a platform for the production of C_4 commodity chemicals

With a family of C_4-selective monofunctional ALDHs in hand, we set out to explore the possibility of producing other important C_4 commodity chemicals from our *n*-butanol pathway. In particular, reduction of 3-hydroxybutyryl-CoA, an intermediate in the *n*-butanol production pathway, yields BDO (Fig. 1A).[28] Upon chemical dehydration, BDOs can be used to produce butadiene for synthetic rubber production, which is currently produced from fossil fuel sources at the level of >10 million metric tonnes per year.[29,30] We therefore set out to screen our 16-member ALDH library for potential candidate enzymes to construct a BDO pathway that can reduce either stereoisomer, (*R*)-3- or (*S*)-3-hydroxybutyryl-CoA (Fig. S7†). In this screen, we found that all candidates were competent to produce 1,3-butanediol at levels from 150–700 mg L^{-1}. Interestingly, we found very little sensitivity to the stereochemistry of the substrate, although preferences between butyryl-CoA and 3-hydroxybutyryl-CoA reduction were observed.

We hypothesized that the differences in *n*-butanol compared to BDO production might arise from limitations in ADH activity for reduction of 3-hydroxybutyraldehyde. As such, we generated an ADH sequence similarity network with the goal of identifying a subfamily with the desired substrate selectivity within the larger superfamily (Fig. S8†).[31,32] A list of candidates within the subfamilies defined by the known C_4-selective ADHs (*bdhA*, *bdhB*, *dhaT*, and *yqhD*) was then generated and screened by co-expression with ALDH46, which showed no stereochemical preference for reduction of 3-hydroxybutyryl-CoA. While several hits were found, it was interesting to note that these ADHs appeared to all be highly specific for the (*R*)-isomer.

During this analysis, we identified HB as a side-product that appears to arise from the reduction of an earlier pathway intermediate, acetoacetyl-CoA (Fig. 1A). HB is also an interesting product as its dehydration produces methyl vinyl ketone,

a reagent used in the production of fine chemicals.[33] Additionally, methyl vinyl ketone is a monomer unit used polymer synthesis.[34] We therefore set out to characterize the selectivity of ALDH-ADH pairs by examining partitioning between BDO and HB (Fig. S9†). This screen indicated HB production is highly specific to the ALDH7-ADH2 pair, providing an even distribution of products at high titer (3.4 ± 0.1 g L^{-1}). On the other end, the ALDH3-ADH22 pair was found to capture a large fraction of the C_4 product pool as BDO (81%), producing 2.9 ± 0.1 g L^{-1} of total products under screening conditions.

A selective pathway for production of HB over BDO was engineered by simply removing the PhaB ketoreductase, forming a truncated pathway by eliminating production of 3-hydroxybutyryl-CoA required for BDO formation (Fig. 1A). With this change, the PhaA-ALDH7-ADH2 pathway generated 2.0 ± 0.2 g L^{-1} HB (Fig. 2C). To selectively produce BDO over HB, we pursued an approach to redirect HB to BDO production by adding a secondary alcohol dehydrogenase (sADH). Specifically, we set out to find a sADH that would catalyze the reduction of HB, resulting from promiscuous ALDH activity on acetoacetyl-CoA, directly to BDO (Fig. S10†).

A number of sADHs have been reported to reduce 4-hydroxy-2-butanone or similar substrates.[35] Several of these sADHs were co-expressed with the ALDH7-ADH2 pair, which consistently produced an even mixture of butanediol and hydroxybutanone. Several of the sADHs enabled a shift in the product profile, producing high levels of BDO (>2 g L^{-1}) and minimal amounts of HB (<250 mg L^{-1}) (Fig. S10†). With these sADHs in hand, we could now control the product profile between HB, BDO, or a mixture of the two (Fig. 2C).

Adaptive evolution of C_4 pathways

With highly specific pathways established for production of *n*-butanol, HB, and BDO in place, we next set out to develop a genetic selection for increasing titers under anaerobic conditions with the long-term goal of gaining new insight into the manipulation of central carbon homeostasis. In contrast to our results with the promiscuous *n*-butanol pathway containing the ethanol short-circuit (Fig. 3A), growth of the fermentation-deficient strain, DH1Δ5, depends solely on *n*-butanol production. Using a set of control plasmids with low, medium, and

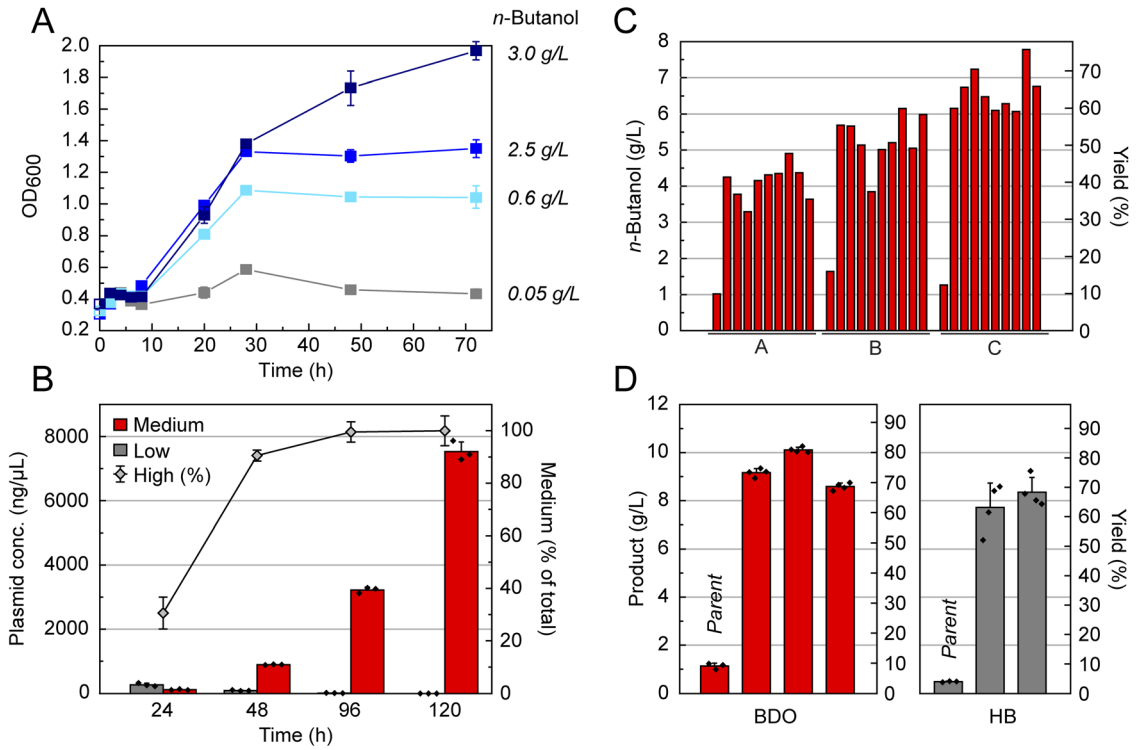

Fig. 3 Development of a genetic selection for evolving C_4 monomer synthesis in *E. coli*. (A) The *n*-butanol pathway complements the deletion of the native fermentation pathways of *E. coli* under anaerobic conditions. *n*-Butanol pathway variants displaying a range of yields were transformed into DH1Δ*5* and cultured anaerobically. Growth was monitored by OD_{600} and *n*-butanol production was quantified at the end of the experiment. All strains were grown in TB with 2.5% (w/v) glucose media. All strains contained pCDF3-ter.aldh46 and one of the following variants of the pBT-Bu2 plasmid in order of increasing butanol production: pBT-0.03HBD, pBT-0.3crt, pTT-Bu2, pBT-Bu2. (B) Enrichment of a medium-producing *n*-butanol strain, DH1Δ*5* pCDF3-ter.aldh46 pBT-0.3crt, was observed when mixed (0.1%) with a low-producing strain DH1Δ5 pCDF3-ter.aldh46 pBT-0.03HBD (99.9%). The mixed culture was grown anaerobically in TB with 2.5% (w/v) glucose media. Changes in the genetic population were monitored using qPCR. (C) A representative adaptive evolution for *n*-butanol production. *E. coli* BW25113Δ*5* pBBR1-aceEF.lpd pT5T33-Bu2 containing either pCWOri.trc-ter-ALDH46.ADH2 [A], pCWOri.trc-ter-ALDH46.ADH8 [B], or pCWOri.trc-ter-ALDH21.ADH2 [C] was subjected to multiple rounds of dilution in M9 containing 10% (v/v) LB and 2.5% (w/v) glucose under anaerobic conditions. Individual clones were then isolated and characterized for their *n*-butanol titers compared to the parent strain. (D) Characterization of BDO and HB strains after adaptive evolution. Data are mean ± s.d. of biological replicates ($n = 3$).

high *n*-butanol productivity, we observe that rescue of DH1Δ5 growth under anaerobic conditions is directly correlated to product titer and thus the capacity of the synthetic pathway to recycle NADH. Indeed, strains complemented with a very low-flux pathway do not grow significantly, if at all, while strains complemented with high flux pathway variants grow similarly to wild-type.

Given the dependence of growth rate on pathway titer, we then tested our ability to enrich cell cultures for high producing variants. To do so, cultures of the low-production strain were seeded with either 0.1% or 1% of the medium-production inoculated strain. Throughout the course of extended anaerobic growth, we observed a significant lag phase dependent on the seeding level (Fig. S11†). In this simulated selection, we tracked *n*-butanol production as well as the abundance of the two different strains using qPCR. In agreement with the growth curves, the abundance of the low-production strain was largely static while the abundance of the medium-production strain was enriched >40-fold as *n*-butanol production initiated (Fig. 3B).

In order to select for variants with improved *n*-butanol productivity under anaerobic conditions, we turned to adaptive evolution after efforts using synthetic mutagenesis methods such as chemical mutagens or UV irradiation appeared to find only local minima in the evolutionary trajectory (Fig. S12†). In this approach, the natural mutation frequency is utilized, which requires longer evolution times but selects for more advantageous mutations and minimizes the occurrence of neutral mutations.[36,37] Since every evolutionary trajectory has the potential to yield different results, we evolved two different host strains, DH1Δ5 and BW25113Δ5, using media ranging in richness from M9, 10% (v/v) LB in M9, and LB, by diluting the culture every 24 h over the course of 4–70 days (Fig. S13†). Using this approach, we obtained evolved strains capable of producing *n*-butanol at levels up to 75% theoretical yield, representing a six-fold improvement over the 12% theoretical yield achieved in the parent strain (Fig. 3C). Although the redox balance is not stoichiometric as it is with *n*-butanol, we were also able to evolve BDO and HB production in DH1Δ5 from 9% to 81% and from 4% to 68% theoretical yield, respectively, in TB (Fig. 3D and S14†). Isolation of pathway plasmids from the evolved strains and transformation into a clean background showed no improvement in product titers, indicating that the relevant mutations were generated on the chromosome (Fig. S14†). Cultures of the evolved strains in rich media in shake-flasks with an oleyl alcohol overlay further yielded titers of up to 47 ± 6 g L^{-1} and >95% yield due to increased *n*-butanol solubility (Fig. S15†). Taken together, the evolved strains demonstrate robust production of a range of C_4 products from acetyl-CoA under anaerobic conditions.

Identifying players in transcriptional re-programming

We took a genome scale approach to explore key factors responsible for the evolution of this large shift in central carbon flow. A total of 31 isolated strains from three independent selections for *n*-butanol (21 strains), BDO (8 strains), and HB (2 strains) production carried out under different growth conditions were sequenced to identify the changes between the genomes of the parental and evolved strains. Interestingly, we found mutations only in a handful of genes, which consistently appeared regardless of selection conditions (Fig. 4A and Table S2†). In addition, a few mutations mapped to the non-coding portions of the genome (0–1 mutation per strain with a total number of 6 distinct mutations from all 31 strains that were sequenced) along with rearrangements that appeared to be mostly associated with mobile elements. Of the mutations in coding regions, the most striking is the finding that poly(A) polymerase (*pcnB*) and/or the RNA polymerase β and β′ subunits (*rpoBC*) were mutated in nearly all of the most successful evolved strains. These two gene loci are involved in regulating the transcriptional landscape of the cell by forming part of the transcription complex (*rpoBC*)[38,39] as well as by controlling the lifetime of mRNAs by polyadenylation (*pcnB*).[40] Mutations in *rne* (ribonuclease E) also occurred frequently (12%) in the evolved *n*-butanol strains.

The discovery that genes involved in RNA metabolism appear to drive metabolic network evolution led us to the hypothesis

A

Gene	Annotation	Pathways
pcnB	Poly(A) polymerase	*n*-Butanol, BDO, HB
rpoB	RNA polymerase β subunit	*n*-Butanol
rpoC	RNA polymerase β′ subunit	*n*-Butanol, BDO
rne	Ribonuclease E	*n*-Butanol
pnp	Polyribonucleotide nucleotidyltransferase	*n*-Butanol, BDO
cadB	Cadaverine/Lys antiporter	*n*-Butanol

B

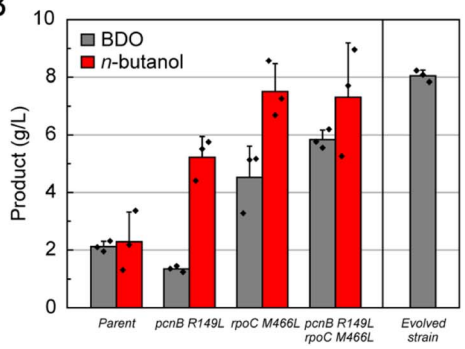

Fig. 4 Characterization of evolved *E. coli* strains for C_4 monomer synthesis. (A) List of genes that were found to be mutated in more than one evolved strain carrying the *n*-butanol, BDO, or HB pathways. (B) Generating the *pcnB* and *rpoC* mutations found in DH1Δ*5*.2406 in a clean genetic background (DH1Δ*5* parent) captures the majority of the improvement observed in the evolved strain, indicating that these two gene loci play an important role in enabling the increases in BDO production. Introduction of the *n*-butanol pathway into DH1Δ*5*.*pcnB*(R149L).*rpoC*(M466L) shows that some aspects of this phenotype can be transferred to other pathways.

that the phenotypic changes were largely being controlled by alterations in the global transcriptional program. This model is consistent with pathway enzymatic activity measurements in cell lysates, which showed no significant increase between parent and evolved strains at the end of a production growth (Fig. S16†). This result suggests that yield increases were not derived from overexpression of heterologous pathway genes. To further characterize this phenomenon, we performed an RNA-Seq experiment on the evolved BDO strain with the largest improvement in production titer (DH1Δ5.2406) containing point mutations in *pcnB* and *rpoC*. We found 126 differentially-expressed genes (β value > 2) between the parental and evolved strain falling into a broad range of categories, including energy production and conversion, amino acid transport and metabolism, cell envelope biogenesis, and carbohydrate transport and metabolism (Fig. S17 and S18†). This large shift in the transcriptome indicates that alterations in acetyl-CoA and central carbon homeostasis may require changes at many metabolic nodes. This data is supported by metabolomics experiment that suggest that acetyl-CoA levels are higher in the evolved strains (Fig. S19†). Genes upregulated in the evolved strain were enriched for the transport GO term, likely to support metabolite uptake and export of BDO (Fig. S20†). Downregulated genes were enriched for the translation, cellular α-amino acid metabolism, nucleotide metabolism, and response to stress GO terms, representing a return to normal carbon homeostasis in the evolved strain (Fig. S20†).

In order to validate the impact of the *pcnB* and *rpoC* mutations, the two mutations observed in this BDO strain (*pcnB* R149L/*rpoC* M466L) were introduced into a clean genetic background. These experiments show that these mutations in *rpoC* and *pcnB* are synergistic, as both are required to achieve a substantive increase in BDO titer compared to the parent (Fig. 4C). Indeed, the double mutant demonstrates a 2.75-fold increase in BDO titers (parent, 2.1 ± 0.1 g L^{-1}; double mutant, 5.8 ± 0.2 g L^{-1}), which recapitulates 73% of the improvement observed in the fully evolved strain (8.1 ± 0.1 g L^{-1}). We were also interested in the generality of these mutations and thus tested their ability to increase the yields of other synthetic pathways. When the *n*-butanol pathway is introduced into the double mutant, we observe a 3.2-fold increase in product titer from 2.3 ± 0.6 to 7.3 ± 1.1 g L^{-1} (Fig. 4C). Altogether, these data show mutations in only two genes, *pcnB* and *rpoC*, are capable of driving a large shift in central carbon metabolism that can be generalized to related pathways utilizing the acetyl-CoA building block.

Conclusions

In this work we have demonstrated the anaerobic production of three industrially relevant C_4 chemicals BDO, HB, and *n*-butanol, at up to >95% of theoretical yield. Furthermore, BDO, HB and *n*-butanol serve as bioproduct precursors to 1,3-butadiene, methyl vinyl ketone, and 1-butene respectively *via* dehydration. Overall, six C_4 chemicals can be accessed from glucose using this single platform. Notably, the production of advanced products from acetyl-CoA can be difficult when switching to anaerobic metabolism. Indeed, acetyl-CoA represents a metabolic checkpoint where carbon is differentiated from pyruvate towards different cell fates, including lowered cell growth and TCA flux, which is one of the major consumers of acetyl-CoA. As such, rational engineering of central carbon pathways for the purpose of re-routing flux to a synthetic product can be quite challenging as it opposes the cell's evolutionary impetus to direct carbon towards biomass. However, using a selection design in which fitness is driven by product titer, strains were identified with up to 5-fold improvements in yield and near quantitative production from the acetyl-CoA building block.

In our system, we found that use of natural adaptive evolution allowed us to rapidly reach high production strains compared to the use of mutagens that increase the mutation rate but appeared to only find local minima in the evolutionary trajectory. Genome-level characterization of these strains revealed that mutations in two gene loci, *pcnB* and *rpoBC*, were sufficient to enable the needed shifts in carbon flow. Interestingly, these mutations have been previously observed in studies of *E. coli* evolution for both growth in M9 (*rpoBC*)[41] and production of *n*-butanol (*pcnB*).[13] Physiological studies suggest that this effect may rely on remodeling the transcriptome by influencing RNA metabolism along with *rne*. Interestingly, a wide range of mutations were identified within these three genes, some of which have been found to be important for activity in biochemical studies.[40] This finding suggests that these four genes could be used for diversity generation at the phenotypic level by inducing pleiotropic changes in the transcriptional landscape. Furthermore, mutations found in the evolved BDO strain could be translated to significant increases in *n*-butanol yields, indicating that these strains could be relevant to the production of other acetyl-CoA-derived products such as fatty acids, polyketides, and isoprenoids.

In conclusion, living systems offer a unique advantage for chemical synthesis to increase product yields through evolution. In particular, central carbon metabolism plays an essential role in cell fitness and thus represents a key regulator and reporter of cellular state.[42] These pathways are subject to tight homeostasis with multiple mechanisms to ensure robustness and reduce sensitivity to change.[43] In this regard, engineered pathways provide an interesting platform where product titer can be treated as a synthetic phenotype or marker for quantitative assessment of genetic traits that lead to large shifts in the regulatory and metabolic network.[44] By using evolution to solve difficult design challenges, we can take advantage of synthetic pathways to identify new strategies to alter behaviours that are hard-wired into the systems-level organization of the host.

Data availability

The datasets supporting this article have been uploaded as part of the ESI.†

Author contributions

MAD and VYY carried out investigation. BF assisted with RNA sequencing experiments. MW carried out experiments for

characterization of aerobic and anaerobic *n*-butanol production in the parent strain. EJK assisted with data analysis and writing of the manuscript. JS designed AdhE2 variants *via* phylogenetic analysis of ALDHs and AdhEs. CAB and DKN were responsible for metabolomic analysis. MAD, VYY, and MCYC conceived of studies. MCYC managed and administered the project. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was funded by the generous support of the National Science Foundation through a CAREER Award (029504-003) to MCYC and the Center for Sustainable Polymers, a National Science Foundation-supported center for Chemical Innovation (CHE-1901635).

References

1 H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A.-L. Barabási, *Nature*, 2000, **407**, 651–654.
2 T. Yamada and P. Bork, *Nat. Rev. Mol. Cell Biol.*, 2009, **10**, 791–803.
3 A. Bordbar, J. M. Monk, Z. A. King and B. O. Palsson, *Nat. Rev. Genet.*, 2014, **15**, 107–120.
4 J. Nielsen and J. D. Keasling, *Cell*, 2016, **164**, 1185–1197.
5 A. L. Meadows, K. M. Hawkins, Y. Tsegaye, E. Antipov, Y. Kim, L. Raetz, R. H. Dahl, A. Tai, T. Mahatdejkul-Meadows, L. Xu, L. Zhao, M. S. Dasika, A. Murarka, J. Lenihan, D. Eng, J. S. Leng, C.-L. Liu, J. W. Wenger, H. Jiang, L. Chao, P. Westfall, J. Lai, S. Ganesan, P. Jackson, R. Mans, D. Platt, C. D. Reeves, P. R. Saija, G. Wichmann, V. F. Holmes, K. Benjamin, P. W. Hill, T. S. Gardner and A. E. Tsong, *Nature*, 2016, **537**, 694–697.
6 I. W. Bogorad, T.-S. Lin and J. C. Liao, *Nature*, 2013, **502**, 693–697.
7 P. Xu, K. Qiao, W. S. Ahn and G. Stephanopoulos, *Proc. Natl. Acad. Sci. U.S.A.*, 2016, **113**, 10848–10853.
8 J. W. Lee, D. Na, J. M. Park, J. Lee, S. Choi and S. Y. Lee, *Nat. Chem. Biol.*, 2012, **8**, 536–546.
9 V. Chubukov, A. Mukhopadhyay, C. J. Petzold, J. D. Keasling and H. G. Martín, *npj Syst. Biol. Appl.*, 2016, **2**, 1–11.
10 K. Jantama, M. j. Haupt, S. A. Svoronos, X. Zhang, J. c. Moore, K. t. Shanmugam and L. o. Ingram, *Biotechnol. Bioeng.*, 2008, **99**, 1140–1153.
11 B. B. Bond-Watts, R. J. Bellerose and M. C. Y. Chang, *Nat. Chem. Biol.*, 2011, **7**, 222–227.
12 J. C. Liao, L. Mi, S. Pontrelli and S. Luo, *Nat. Rev. Microbiol.*, 2016, **14**, 288–304.
13 S. Pontrelli, R. C. B. Fricke, S. S. M. Sakurai, S. P. Putri, S. Fitz-Gibbon, M. Chung, H.-Y. Wu, Y.-J. Chen, M. Pellegrini, E. Fukusaki and J. C. Liao, *Metab. Eng.*, 2018, **49**, 153–163.
14 N. L. Morrow, *Environ. Health Perspect.*, 1990, **86**, 7–8.
15 *Industrial biorenewables: a practical viewpoint*, ed., P. Domínguez de María, Wiley, Hoboken, New Jersey, 2016.
16 L. M. Madeira and M. F. Portela, *Catal. Rev.*, 2002, **44**, 247–286.
17 A. E. Humphrey and S. E. Lee, in *Riegel's Handbook of Industrial Chemistry*, ed. J. A. Kent, Springer Netherlands, Dordrecht, 1992, pp. 916–986.
18 D. L. Nelson, A. L. Lehninger and M. M. Cox, *Lehninger Principles of Biochemistry*, Macmillan, 2008.
19 J. P. van Dijken and W. A. Scheffers, *FEMS Microbiol. Rev.*, 1986, **1**, 199–224.
20 L. Shi and B. P. Tu, *Curr. Opin. Cell Biol.*, 2015, **33**, 125–131.
21 L. Galdieri, T. Zhang, D. Rogerson, R. Lleshi and A. Vancura, *Eukaryotic Cell*, 2014, **13**, 1472–1483.
22 D. P. Clark, *FEMS Microbiol. Rev.*, 1989, **5**, 223–234.
23 S. Y. Lee, J. H. Park, S. H. Jang, L. K. Nielsen, J. Kim and K. S. Jung, *Biotechnol. Bioeng.*, 2008, **101**, 209–228.
24 R. Gheshlaghi, J. M. Scharer, M. Moo-Young and C. P. Chou, *Biotechnol. Adv.*, 2009, **27**, 764–781.
25 J. J. Kim, M. Wang and R. Paschke, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 7523–7527.
26 R. P. McAndrew, Y. Wang, A.-W. Mohsen, M. He, J. Vockley and J.-J. P. Kim, *J. Biol. Chem.*, 2008, **283**, 9435–9443.
27 R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. L. Sonnhammer, J. Tate and M. Punta, *Nucleic Acids Res.*, 2014, **42**, D222–D230.
28 N. Kataoka, A. S. Vangnai, T. Tajima, Y. Nakashimada and J. Kato, *J. Biosci. Bioeng.*, 2013, **115**, 475–480.
29 D. K. Schneiderman and M. A. Hillmyer, *Macromolecules*, 2017, **50**, 3733–3749.
30 *Global 1-Butene Market By Application (LLDPE, HDPE, Valeraldehyde, and Others), By Region (North America, Europe, Asia-Pacific, and Rest of the World) – Demand-Supply, Price, Analysis and Forecast To 2021*, 2016.
31 H. J. Atkinson, J. H. Morris, T. E. Ferrin and P. C. Babbitt, *PLoS One*, 2009, **4**, e4345.
32 S. Zhao, R. Kumar, A. Sakai, M. W. Vetting, B. M. Wood, S. Brown, J. B. Bonanno, B. S. Hillerich, R. D. Seidel, P. C. Babbitt, S. C. Almo, J. V. Sweedler, J. A. Gerlt, J. E. Cronan and M. P. Jacobson, *Nature*, 2013, **502**, 698–702.
33 World Intellectual Property Organization, *WO Pat.* 2011087962A1, 2011.
34 C. S. Marvel and C. L. Levesque, *J. Am. Chem. Soc.*, 1938, **60**, 280–284.
35 A. Matsuyama, Y. Kobayashi and H. Ohnishi, *Biosci., Biotechnol., Biochem.*, 1993, **57**, 348–349.
36 M. Dragosits and D. Mattanovich, *Microb. Cell Factories*, 2013, **12**, 64.
37 J. D. Winkler and K. C. Kao, *Genomics*, 2014, **104**, 406–411.
38 Y. Zuo and T. A. Steitz, *Mol. Cell*, 2015, **58**, 534–540.
39 K. S. Murakami, *J. Biol. Chem.*, 2013, **288**, 9126–9134.
40 Y. Toh, D. Takeshita, T. Nagaike, T. Numata and K. Tomita, *Structure*, 2011, **19**, 232–243.

41 T. M. Conrad, M. Frazier, A. R. Joyce, B.-K. Cho, E. M. Knight, N. E. Lewis, R. Landick and B. Ø. Palsson, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 20500–20505.
42 C. M. Metallo and M. G. Vander Heiden, *Mol. Cell*, 2013, **49**, 388–398.
43 R. A. Cairns, I. S. Harris and T. W. Mak, *Nat. Rev. Cancer*, 2011, **11**, 85–95.
44 H. Alper, J. Moxley, E. Nevoigt, G. R. Fink and G. Stephanopoulos, *Science*, 2006, **314**, 1565–1568.