Quantum chemistry reveals thermodynamic principles of redox biochemistry

Adrian Jinich[1,2], **Avi Flamholz**[3], **Haniu Ren**[1], **Sung-Jin Kim**[4,5], **Benjamin Sanchez-Lengeling**[1], **Charles A. R. Cotton**[6], **Elad Noor**[7], **Alán Aspuru-Guzik**[8,9,10], **Arren Bar-Even**[6]*

1 Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts, United States of America, **2** Division of Infectious Diseases, Weill Department of Medicine, Weill-Cornell Medical College, New York, New York, United States of America, **3** Department of Molecular and Cellular Biology, University of California, Berkeley, Berkeley, California, United States of America, **4** Center for Systems Biology, Massachusetts General Hospital, Boston, Massachusetts, United States of America, **5** Department of Biomedical Engineering, Worcester Polytechnic Institute, Worcester, Massachusetts, United States of America, **6** Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany, **7** Institute of Molecular Systems Biology, ETH Zurich, Zürich, Switzerland, **8** Department of Chemistry and Department of Computer Science, University of Toronto, Ontario, Canada, **9** Vector Institute, Toronto, Ontario, Canada, **10** Biologically-Inspired Solar Energy Program, Canadian Institute for Advanced Research (CIFAR), Toronto, Ontario, Canada

* Bar-Even@mpimp-golm.mpg.de

Abstract

Thermodynamics dictates the structure and function of metabolism. Redox reactions drive cellular energy and material flow. Hence, accurately quantifying the thermodynamics of redox reactions should reveal design principles that shape cellular metabolism. However, only few redox potentials have been measured, and mostly with inconsistent experimental setups. Here, we develop a quantum chemistry approach to calculate redox potentials of biochemical reactions and demonstrate our method predicts experimentally measured potentials with unparalleled accuracy. We then calculate the potentials of all redox pairs that can be generated from biochemically relevant compounds and highlight fundamental trends in redox biochemistry. We further address the question of why NAD/NADP are used as primary electron carriers, demonstrating how their physiological potential range fits the reactions of central metabolism and minimizes the concentration of reactive carbonyls. The use of quantum chemistry can revolutionize our understanding of biochemical phenomena by enabling fast and accurate calculation of thermodynamic values.

Author summary

Redox reactions define the energetic constraints within which life can exist. However, measurements of reduction potentials are scarce and unstandardized, and current prediction methods fall short of desired accuracy and coverage. Here, we harness quantum chemistry tools to enable the high-throughput prediction of reduction potentials with unparalleled accuracy. We calculate the reduction potentials of all redox pairs that can be generated using known biochemical compounds. This high-resolution dataset enables us to uncover global

trends in metabolism, including the differences between and within oxidoreductase groups. We further demonstrate that the redox potential of NAD(P) optimally satisfies two constraints: reversibly reducing and oxidizing the vast majority of redox reactions in central metabolism while keeping the concentration of reactive carbonyl intermediates in check.

Introduction

In order to understand life we need to understand the forces that support and constrain it. Thermodynamics provides the fundamental constraints that shape metabolism [1–5]. Redox reactions constitute the primary metabolic pillars that support life. Life itself can be viewed as an electron transport process that conserves and dissipates energy in order to generate and maintain a heritable local order [6]. Indeed, almost 40% of all known metabolic reactions are redox reactions [7,8]. Redox biochemistry has shaped the study of diverse fields in biology, including origin-of-life [9], circadian clocks [10], carbon-fixation [11], cellular aging [12], and host-pathogen interactions [13]. Previous work has demonstrated that a quantitative understanding of the thermodynamic parameters governing redox reactions reveals design principles of metabolic pathways. For example, the unfavorable nature of carboxyl reduction and carboxylation explains to a large degree the ATP investment required to support carbon fixation [1].

Developing a deep understanding of redox biochemistry requires a comprehensive and accurate set of reduction potential values covering a broad range of reaction types. However, only ~100 reduction potentials can be inferred from experimental data, and these suffer from inconsistencies in experimental setup and conditions. Alternatively, group contribution methods (GCM) can be used to predict a large set of Gibbs energies of formation and reduction potentials [14]. However, the accuracy of this approach is limited, as GCM do not account for interactions between functional groups within a single molecule and GCM predictions are limited to metabolites with functional groups spanned by the model and experimental data.

Quantum chemistry is an alternative modeling approach that has been used to predict redox potentials in the context of numerous applications, such as redox flow batteries, optoelectronics, and design of redox agents [15–27]. Unlike GCM, whose smallest distinct unit is a functional group, quantum chemistry directly relates to the atomic and electronic configuration of a molecule, enabling ab initio prediction of molecular energetics. Here, we adopt a quantum chemistry modeling approach from the field of redox flow battery design [25,26,28] to predict the reduction potentials of biochemical redox pairs. Our approach combines *ab initio* quantum chemistry estimates with (minimal) calibration against available experimental data. We show that the quantum chemical method can predict experimentally derived reduction potentials with considerably higher accuracy than GCM when calibrated with only two parameters. We use this method to estimate the reduction potentials of all possible redox pairs that can be generated from the KEGG database of biochemical compounds [7,8]. This enables us to decipher general trends between and within groups of oxidoreductase reactions, which highlight design principles encoded in cellular metabolism. We specifically focus on explaining the central role of NAD(P) as electron carrier from the perspective of the redox reactions it supports and the role it plays in lowering the concentration of reactive carbonyls.

Results

Quantum chemical predictions of biochemical redox potentials

To facilitate our analysis we divided redox reactions into several generalized oxidoreductase groups which together cover the vast majority of redox transformations within cellular
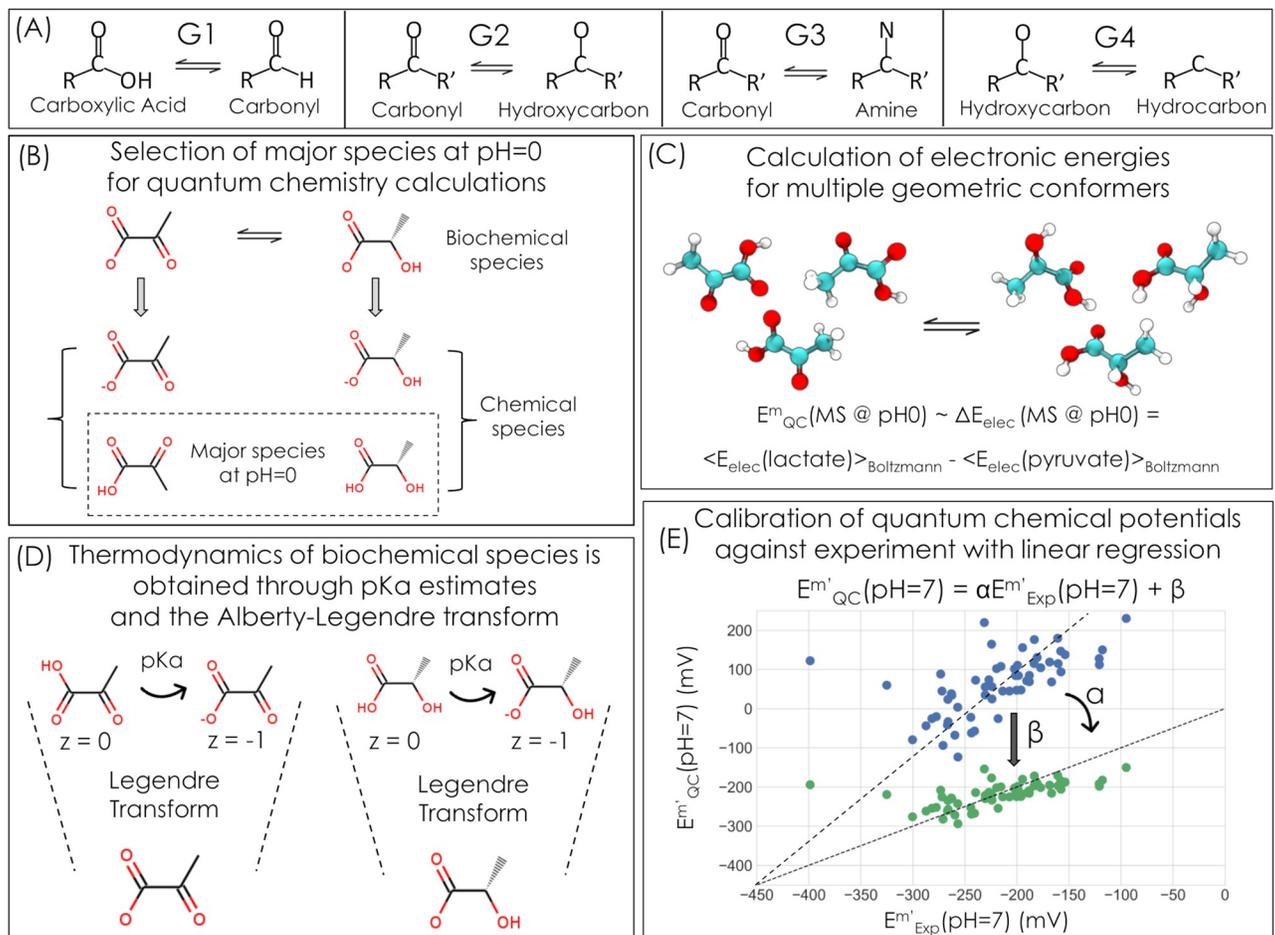
Fig 1. Our study is based on predicting biochemical standard redox potentials using a calibrated quantum chemistry strategy. (A) The four different redox reaction categories considered here are reduction of a carboxylic acid to a carbonyl—G1, reduction of a carbonyl to a hydroxycarbon—G2, or an amine—G3, and reduction of a hydroxycarbon to a hydrocarbon—G4. (B) For each redox reaction of interest, such as reduction of pyruvate to lactate, we select the most abundant protonation state at acidic pH (pH = 0) for quantum chemical simulation. (C) We estimate the chemical redox potential as the difference between Boltzmann-averaged electronic energies of geometric conformers of products and substrates. (D) In order to convert chemical redox potentials to biochemical potentials at pH = 7, we use cheminformatic pKa estimates and the Alberty-Legendre Transform (Supplementary Information). (E) Finally, we use a set of 105 experimental values obtained from the NIST Thermodynamics of Enzyme-Catalyzed Reactions database (TECRDB) [30] and a set of Gibbs formation energies compiled by Robert Alberty [31] (Supplementary Information) to calibrate redox potentials using linear regression.

https://doi.org/10.1371/journal.pcbi.1006471.g001

metabolism (Fig 1A): (G1) reduction of an unmodified carboxylic acid (-COO) or an activated carboxylic acid–i.e., phosphoanhydride (-COOPO$_3$) or thioester (-COS-CoA)–to a carbonyl (-C = O); (G2) reduction of a carbonyl to a hydroxycarbon (-COH, i.e., alcohol); (G3) reduction of a carbonyl to an amine (-CNH$_3$); and (G4) reduction of a hydroxycarbon to a hydrocarbon (-C-C-), which usually occurs via an ethylene intermediate (-C = C-). We note that this categorization corresponds to the treatment of carbon oxidation levels in standard organic chemistry textbooks [29].

We developed a quantum chemistry method for predicting the standard transformed redox potential of biochemical redox reactions. We explored a range of different model chemistries, including combinations of DFT (density functional theory) functionals or wave-function electronic structure methods, basis sets, choice of implicit solvent, and choice of dispersion correction. We found that a DFT approach that uses the double-hybrid functional B2PLYP [32,33] gave the highest prediction accuracy (see Methods for detailed model chemistry description;
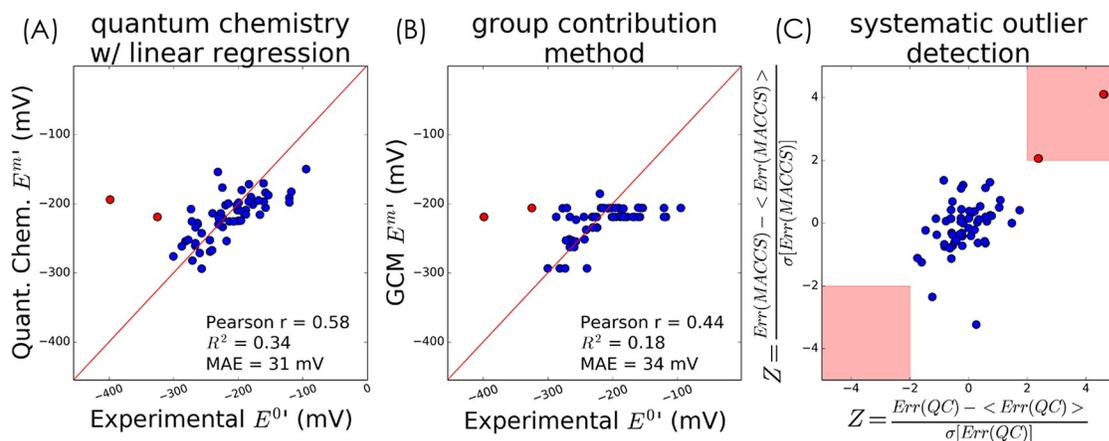
Fig 2. Quantum chemistry model predicts experimentally measured reduction potential with high accuracy. Data shown corresponds to reactions where carbonyls are reduced to hydroxycarbons (group G2). (A) Quantum chemical predictions after calibration (linear regression with 2-parameters); S2 Fig shows how the calibration improves accuracy. (B) Prediction using group contribution method as implemented in eQuilibrator [36,37] (see Methods) (10 parameters for the G2 category) (C) Scatter plot of normalized prediction errors (z-scores) of G2 reactions for molecular fingerprints and quantum chemistry. The indolelactate dehydrogenase (EC 1.1.1.110) and the succinate semialdehyde reductase (EC 1.1.1.61) reactions (red points) have potentially erroneous experimental values.

other model chemistries also gave high accuracy as discussed in the Supplementary Information and S1 Fig). As each biochemical compound represents an ensemble of different chemical species–each at a different protonation state [31] –we applied the following pipeline to predict E'^{m} (Fig 1, see also Methods): (i) a quantum chemical simulation was used to obtain the electronic energies of the most abundant chemical species at pH 0; (ii) we then calculated the difference in electronic energies $\Delta E_{Electronic}$ between the product and substrate of a redox pair at pH 0, thus obtaining estimates of the standard redox potential, E^o; (iii) next, we employed empirical pKa estimates to calculate the energetics of the deprotonated chemical species and used the extended Debye-Huckel equation and the Alberty-Legendre transform [31] to convert E^o to the standard transformed redox potential E'^{m} at pH = 7 and ionic strength I = 0.25 M (as recommended [34]), where reactant concentrations are standardized to 1 mM to better approximate the physiological concentrations of metabolites [1,35]. Finally, (iv) to correct for systematic errors, the predicted E'^{m} values, of each oxidoreductase group, were calibrated by linear regression (two-parameter calibration) against a set of 105 experimentally measured potentials obtained from the NIST Thermodynamics of Enzyme-Catalyzed Reactions database (TECRDB) [30] and the Gibbs formation energy dataset of Robert Alberty [31] (Supplementary Information). We note that we observe empirically that the difference in electronic energies $\Delta E_{Electronic}$ is strongly correlated with the Gibbs reaction energy ΔG_r for these redox systems (S5 Fig) and so we estimate redox potentials using the former in order to reduce computational cost (see SI for details). We also note that the two-parameter calibration is needed mainly since we ignore vibrational enthalpies and entropies of the compounds (Supplementary Information).

As exemplified in Fig 2A and 2B and S2 Fig, the calibration by linear regression significantly improves the accuracy of our quantum chemistry predictions. As shown in Table 1, the predictions of quantum chemistry have a lower mean absolute error (MAE) than those of GCM for all reaction categories. (GCM has a higher Pearson correlation coefficient for category G1, but this is an artifact introduced by a single outlier value, S3 Fig). The improved accuracy is especially noteworthy as our quantum chemical approach derives reduction potentials from first

Table 1. Prediction accuracy of the quantum chemistry and group contribution method modeling approaches.

	G1 (n = 8) Carboxylic Acid to Carbonyl	G2 (n = 59) Carbonyl to Hydroxycarbon	G3 (n = 23) Carbonyl to Amine	G4 (n = 15) Hydroxycarbon to Hydrocarbon
Quantum Chemistry	MAE = 45 mV Pearson r = 0.43 R^2 = 0.19 No. params. = 2	MAE = 31 mV Pearson r = 0.59 R^2 = 0.35 No. params. = 2	MAE = 17 mV Pearson r = 0.70 R^2 = 0.49 No. params. = 2	MAE = 34 mV Pearson r = 0.45 R^2 = 0.21 No. params. = 2
Group Contribution Method	MAE = 52 mV Pearson r = 0.54 R^2 = 0.17 No. params. = 6	MAE = 34 mV Pearson r = 0.48 R^2 = 0.21 No. params. = 13	MAE = 31 mV Pearson r = 0.22 R^2 = -0.23 No. params. = 5	MAE = 66 mV Pearson r = 0.16 R^2 = -3.39 No. params. = 6

The number of available experimental values for each reaction category is indicated in parentheses. MAE = Mean Absolute Error; R^2 = coefficient of determination. Note that for the G1 category, quantum chemistry has a lower MAE, but GCM has higher values of Pearson r. While the Pearson r can range from -1 to 1, R^2 can take on any negative value. A prediction method with the same accuracy as the mean predictor (a constant model that always predicts the mean value of the experimental data) has a value of R^2 = 0; negative values of R^2 indicate prediction accuracies that are worse than the mean predictor. GCM estimates of standard redox potentials were obtained from the implementation by Noor et al. [36,37] used by eQuilibrator (see Methods).

https://doi.org/10.1371/journal.pcbi.1006471.t001

principles and requires only two calibration parameters per oxidoreductase group (α and β in Fig 1E), as compared to GCM which uses 5–13 parameters while achieving lower prediction accuracy (Table 1). Therefore, our quantum chemistry approach can be extended to predict reduction potentials for a wide domain of redox reactions since it does not depend as heavily on empirical measurements. While the quantum chemistry method is computationally more expensive than GCM–with a cost that scales with the number of electrons per molecule (Supplementary Information)–it can still predict the potentials for several hundreds of reactions when run on a typical high-performance computing cluster.

Systematic detection of potentially erroneous experimental values

Inconsistencies between our predictions and experimental measurements can be used to identify potentially erroneous experimental values. However, as such discrepancies might stem from false predictions, we used an independent method to estimate redox potentials. We reasoned that consistent deviation from two very different prediction approaches should be regarded as indicative of potential experimental error. The second prediction approach we used is based on reaction fingerprints [38], where the structure of the reactants involved is encoded as a binary vector (166 parameters without regularization, Supplementary Information). These binary vectors are then used as variables in a regularized regression to correlate structure against a physicochemical property of interest, such as redox potential [38,39]. This approach is similar to the group contribution method (GCM) in that it is based on a structural decomposition of compounds; however, unlike GCM, fingerprints encode a more detailed structural representation of the compounds.

To detect potentially erroneous experimental measurements, we focused on redox potentials of category G2 (carbonyl to hydroxycarbon reduction) as we have abundant experimental information for this oxidoreductase group (see S4 Fig for results with the other categories). As shown in Fig 2C, we normalized the prediction errors by computing their associated z-scores (indicating how many standard deviations a prediction error is from the mean error across all reactions). Two redox reactions stand out as having significantly different experimental and predicted values for both methods (Z>2): indolepyruvate reduction to indolelactate (indolelactate dehydrogenase, EC 1.1.1.110) and succinate semialdehyde reduction to 4-hydroxybutanoate (succinate semialdehyde reductase, 1.1.1.61).

We suggest an explanation for the observed deviation of the first reaction: in the experimental study, the K'$_{eq}$ of indolelactate dehydrogenase was measured using absorbance at 340 nm as an indicator of the concentration of NADH [40]. However, since indolic compounds also have strong absorption at 340 nm [41], this method probably resulted in an overestimation of the concentration of NADH, and thus an underestimation of K'$_{eq}$. Indeed, the experimentally derived E'^{m} is considerably lower (-400 mV) than the predicted one (-190 mV, via quantum chemistry). With regards to the second reaction, succinate semialdehyde reductase, we note that re-measuring its redox potential is of considerable significance as it plays a central role both in carbon fixation–e.g., the 3-hydroxypropionate-4-hydroxybutyrate cycle and the dicarboxylate-4-hydroxybutyrate cycle [11] –as well as in production of key commodities–e.g., biosynthesis of 1,4-butanediol [42].

Comprehensive prediction and analysis of reduction potentials

We used the calibrated quantum chemistry model to predict redox potentials for a database of natural and non-natural redox reactions. We generated this dataset by identifying pairs of metabolites from KEGG [7,8] that fit the chemical transformations associated with each of the four different oxidoreductase groups (Methods). We considered only compounds with fewer than 7 carbon atoms, thus generating a dataset consisting of 652 reactions: 83 reductions of category G1; 205 reductions of category G2; 104 reductions of category G3; and 260 reductions of category G4 (Supplementary Dataset 1). Some of these redox pairs are known to participate in enzyme-catalyzed reactions while others are hypothetical transformations that could potentially be performed by engineered enzymes. We note that our approach to generate reactions is similar to that of the comprehensive Atlas of Biochemistry [43], but we focus solely on the four redox transformations of interest.

Fig 3A shows the distribution of all predicted redox potentials at pH = 7, I = 0.25 M and reactant concentrations of 1 mM, i.e., E'^{m} [14,36]. Fig 3 demonstrates that the value of E'^{m} is directly related to the oxidation state of the functional group being reduced. The general trend is that "the rich get richer" [1,44,45]: more reduced functional groups have a greater tendency to accept electrons, i.e., have higher reduction potentials. Specifically, the reduction potential of hydroxycarbons (G4, $<E'^m> = -15\ mV$) is higher than that of carbonyls ($<E'^m> = -225\ mV$ for both G2 and G3) and the reduction potential of carbonyls is higher than that of unactivated carboxylic acids (G1, $<E'^m> = -550\ mV$). Categories G2 and G3 (reduction of carbonyls to hydroxycarbons or amines, respectively) have very similar potentials because the oxidation state of the functional groups involved is identical (note that this holds for the physiological E'^m but not for E'^o because reactions in the G3 category are balanced with an ammonia molecule as a substrate, thus introducing a factor of RTln(10^{-3}) when converting to the mM standard state). For category G1, activation of carboxylic acids significantly increases their reduction potential (orange line in Fig 3) as the energy released by the hydrolysis of the phosphoanhydride or thioester (~50kJ/mol) activates the reduction: $\Delta E = \frac{50}{nF} \cong 250\ mV$ (n being the number of electrons, F the Faraday constant).

The quantum chemical predictions further enable us to explore detailed structure-energy relationships within each of the general oxidoreductase groups. To exemplify this we focus on the G2 category, as shown in Fig 4. While we find no significant difference between the average E'^{m} of aldehydes and ketones, we can clearly see that the identity of functional groups adjacent to the carbonyl has a significant effect on E'^{m}, as expected. Alpha ketoacids and dicarbonyls have a significantly higher E'^{m} than alpha hydroxy-carbonyls ($\Delta <E'^m> \cong 20\ mV, p < 0.005$) and carbonyls adjacent to hydrocarbons ($\Delta <E'^m> \cong 35\ mV, p < 0.0005$). Carbonyls next to double bonds or aromatic rings have a significantly lower E'^{m} values than alpha hydroxy-carbonyls and carbonyls that are next to hydrocarbons ($\Delta <E'^m> \cong -50\ mV$, and $\Delta <E'^m> \cong$
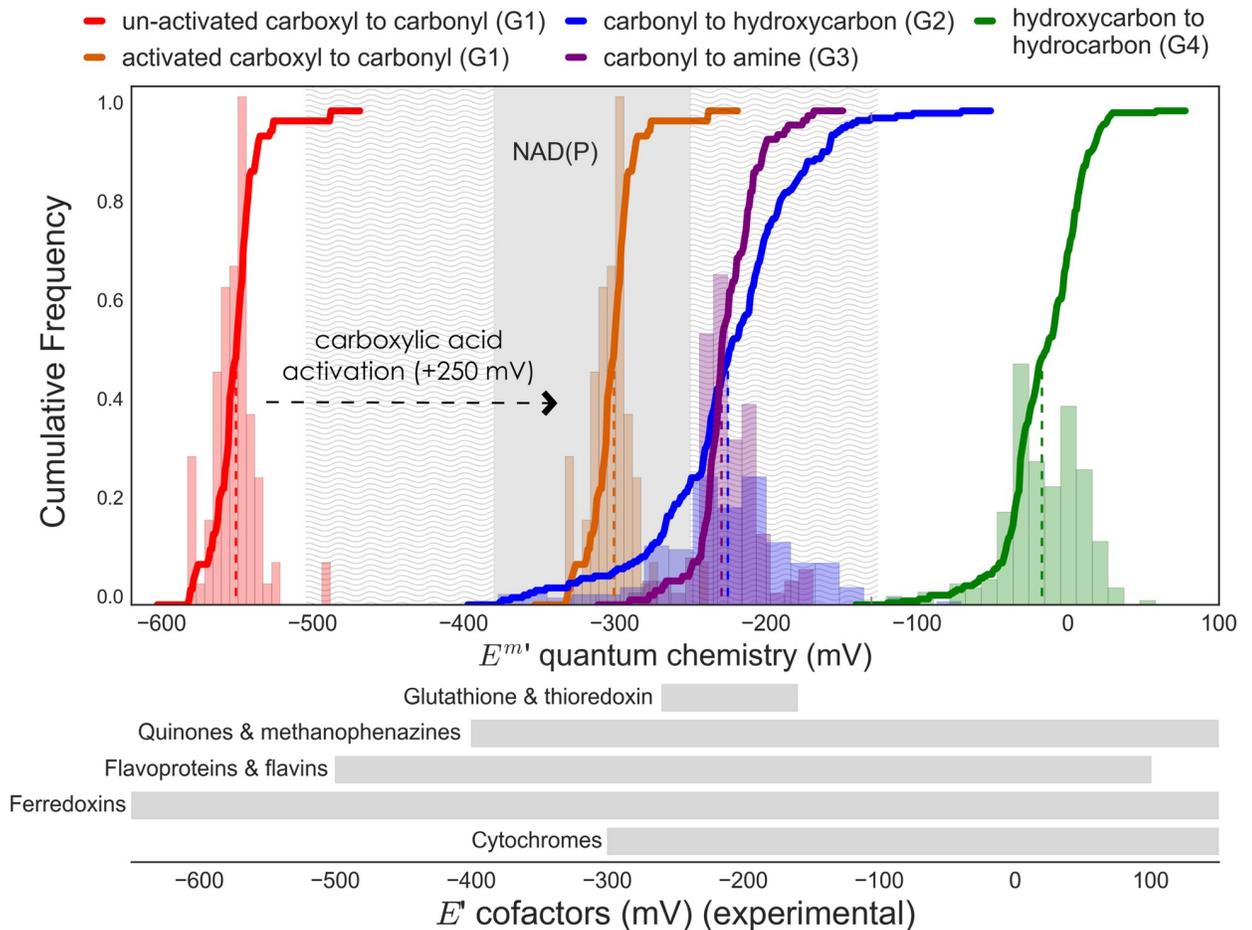
Fig 3. Distributions of predicted standard transformed redox potentials at pH = 7 and I = 0.25 for a dataset of 650 natural and non-natural reactions. The average reduction potentials for each reaction category are (values rounded to nearest multiple of 5): un-activated carboxylic acid to carbonyl (G1: $<E'^m> = -550\ mV$), activated carboxylic acid to carbonyl (activated G1: $<E'^m> = -300\ mV$), carbonyl to hydroxycarbon (G2: $<E'^m> = -225\ mV$), carbonyl to amine (G3: $<E'^m> = -225\ mV$), and hydroxycarbon to hydrocarbon (G4: $<E'^m> = -15\ mV$) Both histograms and cumulative distributions (bold lines, right y-axis) are shown. The distributions for unactivated and activated carboxylic acid to carbonyl reductions (red and purple) are the same, but shifted by +250 mV. Dashed colored lines show the median redox potential for each reaction category. Grey shaded regions corresponds to the range of NAD(P) redox potential, while light grey wavy lines delimit the region of reversible oxidation/reduction by NAD(P)/NAD(P)H. Ranges of reduction potentials for different alternative cofactors are shown as grey rectangles underneath graph (S1 Table).

https://doi.org/10.1371/journal.pcbi.1006471.g003

$-40\ mV$ respectively, $p < 0.0001$). Lactones (cyclic esters), have redox potentials that are significantly lower than any other subgroup within the G2 category. As another validation of the predicted potentials, we found that the reduction potentials of open-chain sugars are significantly higher than those of closed-ring sugars that undergo ring opening upon reduction, where $\Delta <E'^m> \cong 60\ mV$ ($p < 10^{-5}$). This is consistent with the known thermodynamics of closed-ring sugar conformations, e.g., the K_{eq} of arabinose ring opening is ~350[46], which translates to $\Delta E = \frac{RTln(350)}{nF} \cong 75\ mV$, close to the observed average potential difference between the subgroups (R is the gas constant, and T the temperature).

On the biochemical logic of the universal reliance on NAD(P)

While myriad natural electron carriers are known to support cellular redox reactions, NAD(P) has the prime role in almost all organisms, participating in most (>50%) known redox
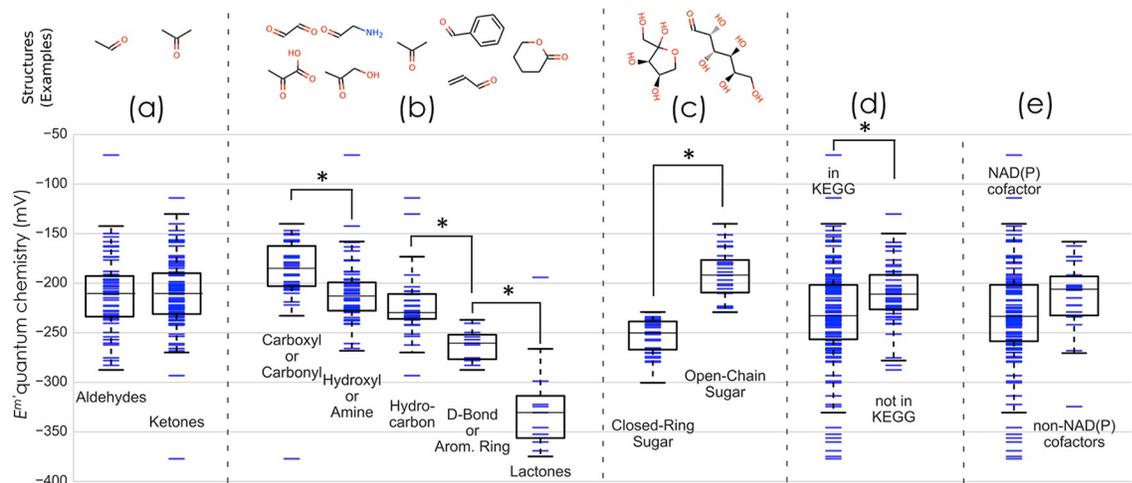
Fig 4. Comparison between the redox potentials of sub-groups for reactions in the G2 category (carbonyl to hydroxycarbon reductions). (A) Aldehydes vs. ketones (non-statistically significant $\Delta <E'^m>$); (B) nearest-neighbor functional group (all subgroups have statistically significant $\Delta <E'^m>$, p<0.005, except hydroxyl/amine and hydrocarbon) (C) closed-ring sugar reduction to open-chain vs. open-chain sugar reduction to open-chain (statistically significant $\Delta <E'^m>$, p<10^{-5}), (D) natural reactions appearing in KEGG vs. non-natural reactions (statistically significant $\Delta <E'^m>$, p<0.005) (E) natural reactions that only use NAD(P) as redox cofactor vs. those that use alternative cofactors (cytochromes, FAD, O_2, or quinones) (non-statistically significant $\Delta <E'^m>$, p = 0.03).

reactions [7,8]. The standard redox potential of NAD(P) is ~ -330 mV (pH = 7, I = 0.25), but as [NADPH]/[NADP] can be higher than 50 and [NADH]/[NAD] can be lower than 1/500, the physiological range of the NAD(P) reduction potential is between -380 mV and -250 mV [35,47–51]. Most cellular redox reactions are therefore constrained to a limited reduction potential range determined by the physicochemical properties and physiological concentrations of NAD(P). By examining the fundamental trends of redox potentials of the different oxidoreductase groups we will show that NAD(P) is well-matched to the redox transformations most commonly found in cellular metabolism.

Fig 3 demonstrates that the reduction potentials of activated acids (activated G1) and carbonyls (G2 and G3) are very similar, such that NAD(P) can support both the oxidation and reduction of nearly all redox couples in these classes. Although the distributions associated with these redox reactions are not entirely contained in the NAD(P) reduction potential range (marked in grey), the reduction potential of a redox pair can be altered by modulating the concentrations of the oxidized and reduced species. As the concentrations of metabolites usually lie between 1 μM and 10 mM [1,4,35,52], the reduction potential of a redox pair can be offset from its standard value by up to $\pm \frac{RT\ln(10^4)}{nF} \cong \pm 120\ mV$ (assuming two electrons are transferred). Therefore, NAD(P) can support reversible redox reactions of compound pairs with E'ᵐ as low as $-380 - 120 = -500\ mV$ and as high as $-250 + 120 = -130\ mV$ (indicated by the light grey regions in Fig 3), a range that encompasses almost all activated acids (activated G1) and carbonyls (G2 and G3 reactions). Outside this range, however, NAD(P)(H) can only be used in one direction of the redox transformation–either oxidation or reduction, but not both. Fig 3 shows that NAD(P)H can support irreversible reductions of hydroxycarbons to hydrocarbons and NAD(P) supports irreversible oxidation of carbonyls to carboxylic acids.

Next, we focus on a small set of redox reactions found in the extended central metabolic network that is shared by almost all organisms: (i) The TCA cycle, operating in the oxidative or reductive direction [53], as a cycle or as a fork [54], being complete or incomplete [54], or with some local bypasses (e.g., [55]); (ii) glycolysis and gluconeogenesis, whether via the EMP
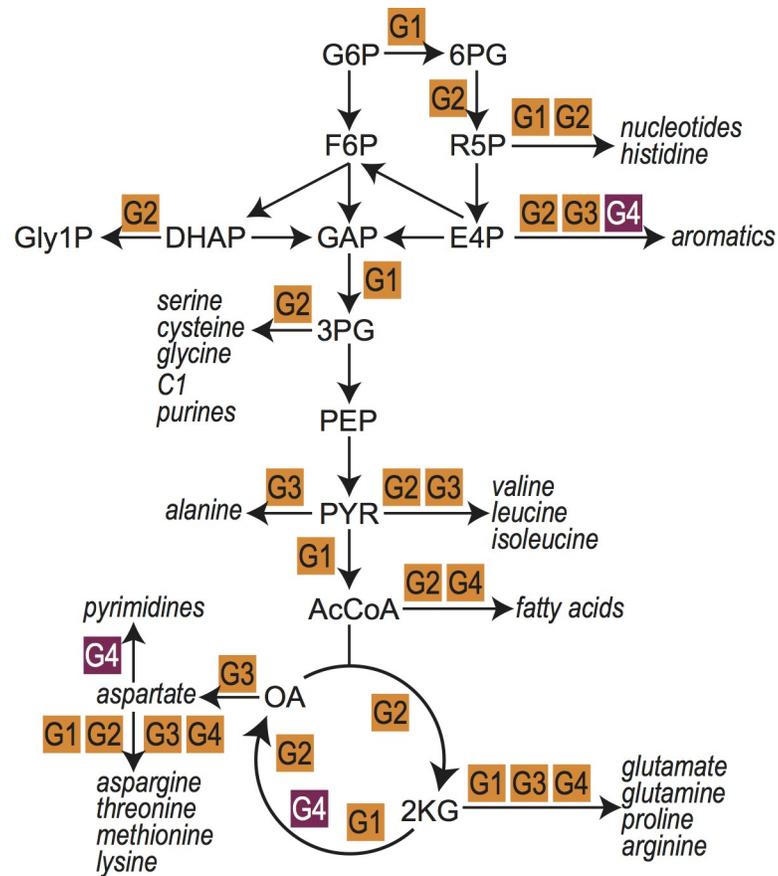
Fig 5. A schematic showing the location of different types of oxidoreductase reactions (oxidoreductase groups 1 to 4) within the extended central metabolic network. We highlight reactions (purple) where a hydrocarbon is oxidized to a hydroxycarbon (G4 reactions, in the direction of oxidation) which generally cannot be sustained by NAD (P) as redox cofactor. See Supplementary Dataset 2 for full set of redox reactions in extended central metabolic network. G6P, Glucose-6-phosphate; F6P, Fructose-6-phosphate; DHAP, Dihydroxyacetone phosphate; GAP, Glyceraldehyde 3-phosphate; Gly1P; Glycerol 1-phosphate; 6PG, 6-Phosphogluconolactone; R5P, Ribulose 5-phosphate; E4P, Erythrose 4-phosphate; 3PG, 3-Phosphoglycerate; PEP, Phosphoenolpyruvate; PYR, Pyruvate; AcCoA, Acetyl coenzyme A; 2KG, 2-Ketoglutaric acid; OA, Oxaloacetate.

or ED pathway [56], having fully, semi or non-phosphorylated intermediates [57]; (iii) the pentose phosphate cycle, working in the oxidative, reductive or neutral direction; and (iv) biosynthesis of amino-acids, nucleobases and fatty acids. As schematically shown in Fig 5, and listed in Supplementary Dataset S2, the ≈ 60 redox reactions that participate in the extended central metabolism almost exclusively belong to one of the following groups: (i) reduction of an activated carboxylic acid to a carbonyl or the reverse reaction oxidizing the carbonyl (9 reactions, G1); (ii) reduction of a carbonyl to a hydroxycarbon or its reverse oxidation (20 reactions, G2); (iii) reduction of a carbonyl to an amine or its reverse oxidation (18 reactions, G3); (iv) irreversible oxidation of carbonyls to un-activated carboxylic acids (5 reactions, G1 in the direction of oxidation); and (v) irreversible reduction of hydroxycarbon to hydrocarbons (4 reactions, G4). Only two central metabolic reactions (marked in magenta background in Fig 5) oxidize hydrocarbons to hydroxycarbons (G4, in the direction of oxidation) and require a reduction potential higher than that of NAD(P): oxidation of succinate to fumarate and oxidation of dihydroorotate to orotate (While formally being oxidation of hydrocarbon to hydroxycarbon, the oxidations of prephenate to 4-hydroxyphenylpyruvate and of arogenate to

tyrosine present a special case since they create a highly stable aromatic ring and hence have enough energy to donate their electrons directly to NAD(P)). Similarly, the extended central metabolic network does not demand the low reduction potential required for the reduction of un-activated carboxylic acids (G1).

The reduction potential range associated with NAD(P) therefore perfectly matches the vast majority of reversible redox reactions in extended central metabolism–i.e., reduction of activated carboxylic acids and reduction of carbonyls (orange, purple and blue distributions in Fig 3)–and can also support the common irreversible redox transformations of extended central metabolism–i.e., reduction of hydroxycarbons and oxidation of carbonyls to un-activated carboxylic acids (green and red distributions in Fig 3). Cells typically rely on secondary redox carriers like quinones and ferredoxins (Fig 3, S1 Table), to support less common reactions, i.e., oxidation of hydrocarbons and reduction of un-activated carboxylic acids.

Why is the reduction potential of NAD(P) lower than the E'^{m} of most carbonyls (Fig 3)? As biosynthesis of an NAD(P) derivative with higher reduction potential presents no major challenge [58], why does this lower potential persist? We suggest that this redox offset plays an important role in reducing the concentrations of cellular carbonyls by making their reduction to hydroxycarbons favorable. It is well known that carbonyls are reactive towards macromolecules, as they spontaneously cross-link proteins, inactivate enzymes and mutagenize DNA [59,60]. As the reduction potential of NAD(P) is lower than most carbonyls, the redox reactions in category G2 (or G3) prefer the direction of reduction, thus ensuring that carbonyls are kept at lower concentrations than their corresponding hydroxycarbons (or amines). Assuming a value of $E' = -330$ mV for NAD(P) and taking the average E'^{m} of the G2 reactions ($<E'^m> \cong -225\ mV$) results in an estimated equilibrium concentration ratio $\frac{[hydroxycarbon]}{[carbonyl]} = exp\left(-\frac{(E[NAD(P)]-<E^m>)nF}{RT}\right) \cong 3500$, thus ensuring very low levels of the carbonyl species. While we do not have many measurements to confirm this prediction, we note one central example: in *E. coli*, the concentration of oxaloacetate is 1–4 μM [61], while the concentration of its conjugated hydroxyacid, malate, is 2–3 mM [52].

For ketoacids and open-ring sugars (which are especially reactive due to the free carbonyl) this effect is even more pronounced as both have especially high reduction potentials (Fig 4). Indeed, the reduction potential of ketoacids is so high that the reverse, oxidative reaction is usually supported by electron donors with a higher potential than NAD(P), for example, quinones, flavins, and even O_2 (e.g., lactate oxidase, glycolate oxidase). Interestingly, the reactions of category G2 that are supported by known enzymes in the KEGG database (75% of reactions in this category) have significantly lower E'^{m} than the remaining reactions, which are not known to be catalyzed by natural enzymes ($\Delta <E'^m> \cong 20\ mV, p < 0.005$). As such, we suggest that the G2 transformations that are known to be enzyme-catalyzed are mainly those that are amenable to redox coupling with NAD(P) (Fig 4D). Within the subset of G2 transformations found in KEGG, those that use redox cofactors other than NAD(P) (such as cytochromes, FAD, O_2, or quinones) have higher E'^{m} values ($\Delta <E'^m> \cong 20\ mV$, not significant $p = 0.03$) than those that use NAD(P) (Fig 4).

Finally, we note that the reduction potential of NADP and activated carboxylic acids (activated G1) overlap almost completely, such that we would not expect NAD(P) to have a strong effect on the ratio between the concentrations of carbonyls and activated acids. This is to be expected as both carbonyls and activated carboxylic acids are reactive–e.g., acetylphosphate and glycerate bisphosphate acetylates proteins spontaneously [62] and acyl-CoA's S-acetylates cellular peptides non-enzymatically [63]. As such, there is no sense in driving the accumulation of carbonyls at the expense of activated carboxylic acids or vice-versa–neither approach would ameliorate non-specific toxicity.

Discussion

In this work, we present a novel approach for predicting the thermodynamics of biochemical redox reactions. Our approach differs radically from group contribution methods, which rely on a large set of arbitrarily-defined functional groups, assume no energetic interactions between groups, and are restricted to metabolites that are decomposable into the groups spanned by the model. In contrast, quantum chemistry directly takes into account the electronic structure of metabolites in solution.

Focusing on specific examples highlights the strengths of our quantum chemical approach as well as various weaknesses of GCM. For example, we find several reactions where the GCM predictions are obviously inaccurate as they are too high to be reasonable: 2-Hydroxy-5-methylquinone ⇔ 2,4,5-Trihydroxytoluene (GCM: $E'^m = 543\ mV$, QC: $E'^o = -158\ mV$); 2-Pyrone-4,6-dicarboxylate ⇔ 2-Hydroxy-2-hydropyrone-4,6-dicarboxylate (GCM: $E'^m = 1406\ mV$, QC: $E'^o = -375\ mV$); and Mevaldate ⇔ (R)-Mevalonate (GCM: $E'^m = 132\ mV$, QC: $E'^o = -190\ mV$). Close inspection of the group matrix underlying these estimates reveals errors in the decomposition of the compounds. Failures in the GCM decomposition are likely due to the complexity of molecular representations in the standard INCHI format [64] and usually occur with aromatic and delocalized electrons. This reflects challenges inherent in group decomposition, which are avoided when using the quantum chemistry approach.

A more illuminating example is that of 3-dehydroshikimate ⇔ shikimate (shikimate dehydrogenase), the sole redox reaction in the shikimate pathway, converting erythrose 4-phosphate and PEP into chorismate. (Chorismate is required for the biosynthesis of aromatic amino-acids, folates, quinones, and important secondary metabolites [65]). GCM predicts a value of $E'^m = -85\ mV$, which, if correct, indicates that the reduction of 3-dehydroshikimate with NAD(P)H is irreversible. On the other hand, quantum chemistry methods predict $E'^m = -268\ mV$, which corresponds to a 6 order-of-magnitude equilibrium concentration difference with respect to the GCM value. The quantum chemistry prediction thus implies reversibility of the oxidoreductase reaction with NAD(P)H. As oxidation of shikimate by NAD(P) has been shown to occur *in-vivo* in gram positive bacteria [66], it is clear that the GCM prediction is wrong and that the quantum chemistry approach provides a more accurate assessment of the thermodynamic potential of this important biochemical reaction.

Unlike previous efforts [67,68], our quantum chemistry approach relies on a two-parameter calibration for each oxidoreductase reaction category, which reduces computational cost by avoiding the need to calculate vibrational enthalpies and entropies (Supplementary Information). In future studies, improvements in accuracy could be achieved by exploring a larger space of quantum model chemistries, or—if more experimental data becomes available—calibrating using more sophisticated regression techniques, such as Gaussian Process regression [69]. Yet, as we have shown, the current procedure is sufficient to yield high coverage and accuracy at a reasonable computational cost.

In contrast with GCM methods, our calibration parameters can be at least partially interpreted. One important contribution to the systematic bias in the raw quantum calculations (i.e. the y-intercept in the linear regression) comes from neglecting the vibrational component of the molecular enthalpy. Interpreting the slope parameter is more complex, yet examples in the literature show that it can be traced back to the choice of solvation [70] or—in the context of modeling quinone derivatives–to the basis set incompleteness and the shortcomings of the DFT exchange correlation functionals [71]. We note, however, that faster computational resources will eventually enable full *ab initio* prediction of hundreds of standard transformed redox potentials, rendering the two-parameter calibration and the use of empirical pKa values obsolete [15,72].

Importantly, the quantum chemical strategy is not subject to the inconsistencies that plague experimental databases. Experimental values are measured in a wide range of different conditions, including temperature, pH, ionic strength, buffers, and electrolytes. In many cases, the exact measurement conditions are not reported, making it practically impossible to account for these factors. Thus, even if we were to gain access to more experimental data, the lack of systematically applied conditions makes such resources problematic. In contrast, quantum chemical simulations can be performed in consistent, well-defined conditions.

Why does the primary biological reduction potential range lie between -370 mV and -250 mV? One possibility is a frozen evolutionary accident. In this view, NAD(P) was available early in evolution and was found useful in supporting multiple redox reactions; as such, it was fixed as the central redox carrier before the Last Universal Common Ancestor (LUCA). While we cannot rule out this explanation, we suggest an alternative: that the primary reduction potential range represents a near optimal adaptation given biochemical constraints and selection pressures imposed throughout evolution. This idea is supported by the fact that most extant electron carriers already existed in LUCA [6], and yet none have as extensive a role in metabolism as NAD(P). Furthermore, derivatives of NAD(P) are simple to synthesize biochemically–e.g. deamino-NAD is a precursor of NAD–and can have considerably shifted reduction potentials [58]. Despite this, no organism has been found to rely on such derivatives. Finally, the deaza-flavin coenzyme F420 is a prominent electron carrier in the central metabolism of methanogens and other prokaryotes [73,74], and has a reduction potential around -340 mV [75], almost identical to that of NAD(P). Hence, even organisms that partially replace NAD(P) use a carrier with a similar reduction potential.

The enhanced resolution provided by quantum chemistry uncovers important patterns not accessible using traditional analyses. Exemplifying this, we found that the main cellular electron carrier, NAD(P), is 'tuned' to reduce the concentration of reactive carbonyls, thereby keeping the cellular environment more chemically stable. Yet, this protection comes at a price: the oxidation of hydroxycarbons is thermodynamically challenging and often requires the use of electron carriers with higher reduction potential. A recent study demonstrates the physiological relevance of this thermodynamic barrier: the NAD-dependent 3-phosphoglycerate dehydrogenase–the first enzyme in the serine biosynthesis route–can sustain high flux in spite of its unfavorable thermodynamics only through coupling with the favorable reduction of 2-ketoglutarate [76].

Our analysis further supports the previous assertion that the TCA cycle has evolved in the reductive direction [53,77]. While all the other electron transfer reactions in the extended central metabolism belongs to oxidoreductase groups that can be supported by NAD(P)(H), oxidation of succinate–a key TCA cycle reaction–cannot be carried by this electron carrier. As the reverse reaction, i.e., fumarate reduction, can be support by NADH [78,79], it is reasonable to speculate that the reaction first evolved in the reductive direction, and only later was adapted to work in the oxidative direction using an alternative cofactor.

So long as sufficient experimental data is available to allow for calibration, our approach can be extended to other types of biochemical reactions. For example, understanding the thermodynamics of carboxylating and decarboxylating enzymes–the "biochemical gateways" connecting the inorganic and the organic world– could pave the way for the identification of highly efficient, thermodynamically favorable carbon fixation pathways based on non-standard but promising reaction chemistries [80,81]. In this way, high-resolution thermodynamic analyses may provide much needed insight for the engineering of microbes to address global challenges.

Methods

We performed quantum chemical simulations—geometry optimizations followed by electronic single point energy (SPE) calculations—on the major species (MS) of each metabolite of interest at pH = 0, which corresponds to the most positively charged species (see below and SI for details on the model chemistries used). Running calculations on these reference protonation states yields estimates for the standard redox potential, E^o(*major species at pH* = 0). Using pKa values from the ChemAxon calculator plugin (Marvin 17.7.0, 2017, ChemAxon)—a cheminformatics software widely used in the field of biochemical thermodynamics [4,14,36,67,82–85] - the extended Debye-Huckel equation and the Alberty-Legendre transform, we converted E^o(*MS at pH* = 0) to the standard (standardized to 1 mM) transformed redox potential of interest, E'^m(*pH* = 7,*I* = 0.25) [31,34]. To correct for systematic errors in both the quantum chemical predictions and the pKa estimates, we calibrate the resulting E'^m(*pH* = 7,*I* = 0.25) values against experimental data using linear regression, performing a separate calibration for each of the four different redox reaction categories. We further detail each of these steps below (see also Supplementary Information).

Quantum chemical geometry optimizations

For each metabolite, we generated ten initial geometric conformations using ChemAxon's calculator plugin (Marvin 17.7.0, 2017, ChemAxon). Quantum chemistry calculations were performed using the Orca software package (version 3.0.3) [86]. Geometry optimizations were carried out using DFT, with the B3LYP functional and Orca's predefined DefBas-2 basis set (see S3 Table for detailed basis set description). The COSMO implicit solvent model [87] was used, with the default parameter values of epsilon = 80.4 and refrac = 1.33. DFT-D3 dispersion correction [88] using Becke-Johnson damping [89] was also included.

Quantum chemical electronic single point energies (SPE) and calibration against experimental values

Single point energy (SPE) calculations yield the value of the electronic energy $E_{Electronic}$ for each conformer at their optimized geometry. We used the optimized geometries obtained using DFT as inputs for SPE calculations (see below and SI for details on the SPE model chemistry selected). Substrate and product conformers were sampled according to a Boltzmann distribution. By taking the difference of products' and substrates' $E_{Electronic}$ values, we obtain $\Delta E_{Electronic}$, which we treat as directly proportional to the standard reduction potential of the major species at pH 0:

$$E^o(MS\ at\ pH = 0) \sim -\frac{\Delta E_{Electronic}}{nF}$$

The use of $\Delta E_{Electronic}$ to approximate the reduction potential as opposed to ΔG_r^o (which includes rotational and vibrational enthalpies and entropies) reduces computational cost and is motivated by the empirical observation that there is a strong correlation between $\Delta E_{Electronic}$ and ΔG_r^o for these redox systems (S5 Fig, see SI for details). We note that we subtracted the energy of molecular hydrogen (obtained with the same SPE model chemistry) from $\Delta E_{Electronic}$ in order to get redox the potentials relative to the standard hydrogen electrode. A similar approach has been used to model redox reactions in the context of organic redox flow batteries [28].

We use cheminformatic pKa estimates (Marvin 17.7.0, 2017, ChemAxon), the extended Debye-Huckel equation and the Alberty-Legendre transform (16, 17) to convert both the experimental standard redox potentials and the quantum chemical predictions of E^o(*MS at pH* = 0) to the *transformed* redox potentials standardized to 1 mM, E'^m(*pH* = 7,*I* = 0.25). Then,

independently for each redox category, we performed linear regressions between the $E'^m(pH = 7, I = 0.25)$ values and the available experimental redox potentials. The calibration via linear regression was implemented using the SciKit learn Python library.

In order to optimize prediction accuracy, we ran geometry optimization and SPE calculations using a large diversity of model chemistries, generated by selecting one of ten possible DFT functionals, two wave function electronic structure methods, three possible basis sets, the option of adding implicit solvation, as well as a correction to account for dispersion interactions (S1 Fig, and see SI for details). Optimizing for Pearson correlation coefficients r, we selected the following model chemistry to predict reactions without experimentally measured potentials: a DFT approach with the double-hybrid functional B2PLYP [32,33], the DefBas-5 Orca basis set (see S3 Table for detailed basis set description), COSMO implicit solvent [87], and D3 dispersion correction [88]. To avoid overfitting, we trained the model chemistry optimization procedure on the experimental data for the G3 reaction category (carbonyl to amine reduction), and validate its accuracy on the rest of the oxidoreductase reaction categories (Table 1 and S3 Fig). Hybrid and double-hybrid DFT functionals have been shown to accurately capture the thermochemistry and noncovalent interactions of molecules when compared with coupled cluster results [90,91]. Therefore, we select this double-hybrid DFT approach covers the relevant physics of our problem while minimizing computational cost and maximizing predictive power. Although we explored a large set of DFT functionals, wave function methods, and basis sets, further improvements could be achieved by exploring a larger space of model chemistries, including the geometry optimization procedure, conformer generation method, as well as explicit solvation models [15]. For example, adapting a recent highly accurate method (tested on four molecules) based on the Linear Response Approximation (LRA) to the large scale prediction of E'^m values would be an interesting direction [72].

Predicting redox potentials with molecular fingerprints and group contribution method

We used the RDKit software tool (http://www.rdkit.org), to obtain binary molecular fingerprints of each compound of interest. Because of the relatively small size of our training sets and in order to minimize overfitting, we used MACCS Key 166 fingerprints instead of other popular Morgan circular fingerprints [92]. We concatenated each redox half-reaction substrate/product fingerprint pair into a single reaction fingerprint [38] and used these as input training data for regularized linear regression. We then performed an independent regularized regression for each of the four different redox reaction categories.

To obtain group contribution estimates of redox potentials, we use the group matrix and the group energies of Noor *et al.* [36] used in eQuilibrator [37], an online thermodynamics calculator. We note that eQuilibrator uses the component contribution method (CCM) which combines group contribution energies with experimental reaction or formation Gibbs energies ("reactant contributions") whenever these are available. That is, for reactions with available experimental data, eQuilibrator will return the experimental energies. Thus, for fair comparison against quantum chemistry we used the GCM code underneath eQuilibrator to obtain the group contribution estimates for all reactions in our test set. Just like the quantum chemical predictions, the GCM estimates were standardized to the $E'^m(pH = 7, I = 0.25)$ state.

Systematic detection of reactions with potentially erroneous experimental values

We design a strategy to detect reactions with potentially erroneous experimental values as listed in the NIST Thermodynamics of Enzyme-Catalyzed Reactions Database (TECRDB)

[30]. We identify reactions whose predicted potential deviates from experiment by a similar amount for both the calibrated quantum chemistry and fingerprint-based modeling approaches. In order to make the errors associated to the two different modeling methods comparable, we normalize the prediction errors by computing their associated z-scores: $Z_{Err} = \frac{(Err-\mu)}{\sigma}$. We set a threshold value for the z-score of Z = 2, such that reactions with $Z_{Err}(QC) > 2$ and $Z_{Err}(fingerprints) > 2$ are assigned a high likelihood of having an erroneously tabulated experimental value in NIST-TECRDB.

Generation of comprehensive database of natural and non-natural redox reactions

To generate a database of all possible redox reactions involving natural compounds, we use a decomposition of all metabolites into functional groups as per the group contribution method [36]. We find pairs of metabolites in the KEGG database with functional group vectors whose difference matches the reaction signature of any of the redox reaction categories of interest. For example, pairs of metabolites in the G1 category will have a group difference vector with a +1/-1 in the element corresponding to a carbonyl/carboxylic acid functional group respectively (see SI for details). We note that every reaction generated by this strategy can be uniquely assigned to one of the four redox categories considered.

Using this method we succeeded in generating a rough database of redox reactions. However, additional manual and semi-automated data cleansing was required to get the final version of the database (see SI for further details). For example, use of the group difference vectors failed to account for the chirality of the metabolites, and in some instances stereochemistry was not maintained throughout the reaction. In order to solve this, we applied an additional filter, which used the conventions for assigning chirality (R/S, L/D) present in molecule names to match chirality between the substrate and product. Sugars proved to be especially problematic as those reactions did not maintain stereochemistry throughout; for these reactions, the above filtering method did not suffice, often keeping incorrect reactions such as L-Xylonate → L-Arabinose. For this, we used molecular naming conventions to eliminate the wrong reactions (see SI for further details).

Statistically significant differences between average E'^{m} values for distinct structural groups

We performed Welch's unequal variance t-test to obtain the p-value for the null hypothesis that pairs of different reaction subcategories within group G2 have identical average E'^{m} values (Fig 4). Welch's t-test is an adaptation of Student's t-test which does not assume equal variances.

Supporting information

S1 Table. The range of range of potentials for the most important redox cofactors in biochemistry. S1 Table shows the physiological range of reduction potentials for the major classes of biological electron carriers, as determined by their physicochemical properties and characteristic intracellular concentrations.
(DOCX)

S2 Table. Linear regression coefficients obtained from calibrating the raw redox potential estimates obtained from the quantum single point energy (SPE) model chemistry. The model chemistry used consists of density functional theory with the B2PLYP double-hybrid functional, the DefBas-5 Orca basis set (see S3 Table for detailed basis set description), the

COSMO implicit solvent, and the D3 dispersion correction.
(DOCX)

S3 Table. A detailed description of the Default Basis (DefBas) sets in Orca version 3.0.3.
The notation SV(xxx/yyy) refers to the SV basis set with polarization functions xxx and diffuse functions yyy.
(DOCX)

S4 Table. Prediction accuracy of the quantum chemistry, molecular fingerprints, and group contribution method modeling approaches. The number of available experimental values for each reaction category is indicated in parentheses. MAE = Mean Absolute Error; R^2 = coefficient of determination. The quantum model chemistry uses the double hybrid functional B2PLYP with the DefBas-5 default Orca basis set (see S3 Table for detailed basis set description), the COSMO implicit solvent, and the D3 dispersion correction. While the Pearson r can range from -1 to 1, R^2 can take on any negative value. A prediction method with the same accuracy as the mean predictor (a constant model that always predicts the mean value of the experimental data) has a value of $R^2 = 0$; negative values of R^2 indicate prediction accuracies that are worse than the mean predictor.
(DOCX)

S1 Fig. Prediction accuracy, as measured using Pearson r coefficient, and average runtimes per molecular conformer for different quantum single point energy (SPE) model chemistries. The accuracy measures is obtained from comparing the predicted $E'^m(pH = 7, I = 0.25)$ values against available experimental data. Data corresponds to prediction accuracy on the G3 reaction category, which consists of reductions of carbonyls to amines. Mean runtime is calculated over all molecular conformers involved in the simulation of the G3 reaction set with available experimental data. As detailed in section 2.7 "Systematic model chemistry exploration to optimize prediction accuracy", the SPE model chemistries were obtained from searching over a subspace of possible model chemistries generated from selecting a DFT (or wave function method), a basis set, an implicit solvent model, and a dispersion correction from a total set of: 10 different DFT functionals and 2 wave-function methods, 3 possible basis sets, the option of adding the Conductor-like Screening Model (COSMO) for implicit solvation, as well as the D3 dispersion correction. See Supplementary Dataset 5 for detailed model chemistry descriptions. The option of including or excluding DFT-D3 dispersion correction in the geometry optimization procedure was also considered.
(TIF)

S2 Fig. Predicting biochemical redox potentials of carbonyl to hydroxycarbon reactions (category G2) with different approaches. (A-C) Calibrating quantum chemical estimates through linear regression (2-parameters per reaction category) significantly improves prediction accuracy. Quantum chemical predictions were performed using the double-hybrid DFT functional B2PLYP, the DefBas-2 Orca basis set, COSMO implicit solvent, and D3 dispersion correction (S1 Text). Points in red correspond to reactions which consistently appear as outliers across modeling approaches: the indolepyruvate reduction to indolelactate and succinate semialdehyde reduction to 4-hydroxybutanoate (D-E) Prediction accuracy of group contribution method (10 parameters for the G2 category) and molecular fingerprints (166 parameters calibrated with regularized Lasso regression). (F) Scatter plot of normalized prediction errors (z-scores) of G2 reactions for molecular fingerprints and quantum chemistry. The indolelactate dehydrogenase (EC 1.1.1.110) and the succinate semialdehyde reductase (EC 1.1.1.61) reactions have potentially erroneous experimental values.
(TIF)

S3 Fig. Scatter plots of experimental redox potentials and predicted potentials with the selected calibrated quantum chemistry approach (upper four panels) and group contribution method (GCM) (lower four panels) for all four redox categories. Quantum chemical predictions were performed using the double-hybrid DFT functional B2PLYP, the DefBas-2 Orca default basis set, the COSMO implicit solvent, and D3 dispersion correction (S1 Text). Data corresponds to experimental values and predictions at the pH = 7 and I = 0.25 biochemical state. G1: reduction of an unmodified carboxylic acid (-COO) to a carbonyl (-C = O); G2: reduction of a carbonyl to a hydroxycarbon (-COH, i.e., alcohol); G3: reduction of a carbonyl to an amine (-CNH3); and G4: reduction of a hydroxycarbon to a hydrocarbon (-C-C-).
(DOCX)

S4 Fig. Detection of experimental outliers using a calibrated quantum chemistry approach and MACCS fingerprint predictions for all four reaction categories.
(TIF)

S5 Fig. Correlation between quantum chemical estimates of $\Delta E_{Electronic}$ and $\Delta G_r'^o$. Each redox reaction category is shown in a different color. G1—reduction of carboxyl to aldehyde; G2—reduction of carbonyl (ketone or aldehyde) to hydroxyl; G3—reduction of carbonyl to amine; G4—reduction of hydroxyl to hydrocarbon. $\Delta E_{Electronic}$ was obtained from single point energy (SPE) calculations, while $\Delta G_r'^o$ is obtained by additionally including rovibrational contributions to Gibbs formation energy.
(TIF)

S6 Fig. Correlation between standard transformed redox potential predictions (pH = 7, I = 0.25) using calibrated quantum chemistry with our top-two model chemistries. As discussed in the S1 Text, the prediction accuracy of the calibrated model chemistries was evaluated using the experimental data for the G3 reaction category only (to avoid overfitting). The labels refer to the quantum model chemistry used to perform a single point energy (SPE) calculation on geometry-optimized conformers. For both SPE model chemistries, geometry optimizations were performed using B3LYP functional, Orca's predefined DefBas-2 basis set (S3 Table), COSMO implicit solvent model and DFT-D3 dispersion correction.
(TIF)

S7 Fig. Cumulative distribution functions of runtimes for geometry optimization and single point energy (SPE) estimates using our quantum chemistry method. Distributions are over the entire set of molecular conformers used in our study. Geometry optimizations were performed out using DFT, with the B3LYP functional and Orca's predefined DefBas-2 basis set, as well as the COSMO implicit solvent model (see SI section 2.3). The cumulative distributions of SPE runtimes are shown for the two best-performing model chemistries: the linear-scaling coupled cluster method DLPNO-CCSD(T), with the DefBas-4 Orca basis set (S3 Table), COSMO, implicit solvent, and D3 dispersion correction; and the double-hybrid functional B2PLYP, the DefBas-5 Orca basis set (S3 Table for detailed description), COSMO implicit solvent, and D3 dispersion correction (see SI section 2.7 for further details).
(TIF)

S1 Text. Supplementary material for "quantum chemistry reveals thermodynamic principles of redox biochemistry".
(DOCX)

S1 Dataset. Contains predicted standard redox potentials (group contribution method and calibrated quantum chemistry) an experimental potentials for all redox pairs considered

in this work.
(XLSX)

S2 Dataset. Contains the full set of redox reactions in the extended central metabolic network.
(XLSX)

S3 Dataset. Contains the full set of compound names, KEGG compound identifiers, smiles strings (for the major species at pH = 0), and charge (for the major species at pH = 0) used in this work.
(XLSX)

S4 Dataset. Contains the structural categorization of compounds in the G2 category used to obtain the structure-energy relationships in Fig 4.
(XLSX)

S5 Dataset. Contains the details of all the model chemistries tested during the optimization procedure.
(XLSX)

S6 Dataset. Contains the raw quantum chemical electronic energies–using a variety of model chemistries—Calculated for up to 10 geometrical conformers of each compound considered in this work.
(XLSX)

Author Contributions

Conceptualization: Adrian Jinich, Avi Flamholz, Elad Noor, Alán Aspuru-Guzik, Arren Bar-Even.

Data curation: Adrian Jinich, Haniu Ren, Sung-Jin Kim.

Formal analysis: Adrian Jinich, Avi Flamholz, Haniu Ren, Sung-Jin Kim, Arren Bar-Even.

Funding acquisition: Alán Aspuru-Guzik.

Investigation: Adrian Jinich, Haniu Ren.

Project administration: Arren Bar-Even.

Resources: Alán Aspuru-Guzik.

Software: Adrian Jinich, Haniu Ren, Sung-Jin Kim, Elad Noor.

Supervision: Alán Aspuru-Guzik, Arren Bar-Even.

Visualization: Benjamin Sanchez-Lengeling, Arren Bar-Even.

Writing – original draft: Adrian Jinich, Haniu Ren, Sung-Jin Kim, Arren Bar-Even.

Writing – review & editing: Adrian Jinich, Avi Flamholz, Benjamin Sanchez-Lengeling, Charles A. R. Cotton, Elad Noor, Alán Aspuru-Guzik, Arren Bar-Even.

References

1. Bar-Even A, Flamholz A, Noor E, Milo R. Thermodynamic constraints shape the structure of carbon fixation pathways. Biochim Biophys Acta. 2012; 1817: 1646–1659. https://doi.org/10.1016/j.bbabio.2012.05.002 PMID: 22609686

2. Bar-Even A. Does acetogenesis really require especially low reduction potential? Biochim Biophys Acta. 2013; 1827: 395–400. https://doi.org/10.1016/j.bbabio.2012.10.007 PMID: 23103387

3. Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R. Glycolytic strategy as a tradeoff between energy yield and protein cost. Proc Natl Acad Sci U S A. 2013; 110: 10039–10044. https://doi.org/10.1073/pnas.1215283110 PMID: 23630264

4. Noor E, Bar-Even A, Flamholz A, Reznik E, Liebermeister W, Milo R. Pathway thermodynamics highlights kinetic obstacles in central metabolism. PLoS Comput Biol. 2014; 10: e1003483. https://doi.org/10.1371/journal.pcbi.1003483 PMID: 24586134

5. Ataman M, Hatzimanikatis V. Heading in the right direction: thermodynamics-based network analysis and pathway engineering. Curr Opin Biotechnol. 2015; 36: 176–182. https://doi.org/10.1016/j.copbio.2015.08.021 PMID: 26360871

6. Schoepp-Cothenet B, van Lis R, Atteia A, Baymann F, Capowiez L, Ducluzeau A-L, et al. On the universal core of bioenergetics. Biochim Biophys Acta. 2013; 1827: 79–93. https://doi.org/10.1016/j.bbabio.2012.09.005 PMID: 22982447

7. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 2008; 36: D480–4. https://doi.org/10.1093/nar/gkm882 PMID: 18077471

8. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28: 27–30. PMID: 10592173

9. Russell MJ, Hall AJ. The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. J Geol Soc London. 1997; 154: 377–402. PMID: 11541234

10. Stangherlin A, Reddy AB. Regulation of circadian clocks by redox homeostasis. J Biol Chem. 2013; 288: 26505–26511. https://doi.org/10.1074/jbc.R113.457564 PMID: 23861436

11. Bar-Even A, Noor E, Milo R. A survey of carbon fixation pathways through a quantitative lens. J Exp Bot. 2012; 63: 2325–2342. https://doi.org/10.1093/jxb/err417 PMID: 22200662

12. Sohal RS, Orr WC. The redox stress hypothesis of aging. Free Radic Biol Med. 2012; 52: 539–555. https://doi.org/10.1016/j.freeradbiomed.2011.10.445 PMID: 22080087

13. Kumar A, Farhana A, Guidry L, Saini V, Hondalus M, Steyn AJC. Redox homeostasis in mycobacteria: the key to tuberculosis control? Expert Rev Mol Med. 2011; 13: e39. https://doi.org/10.1017/S1462399411002079 PMID: 22172201

14. Noor E, Bar-Even A, Flamholz A, Lubling Y, Davidi D, Milo R. An integrated open framework for thermodynamics of reactions that combines accuracy and coverage. Bioinformatics. 2012; 28: 2037–2044. https://doi.org/10.1093/bioinformatics/bts317 PMID: 22645166

15. Marenich AV, Ho J, Coote ML, Cramer CJ, Truhlar DG. Computational electrochemistry: prediction of liquid-phase reduction potentials. Phys Chem Chem Phys. 2014; 16: 15068–15106. https://doi.org/10.1039/c4cp01572j PMID: 24958074

16. Ho J, Coote M, Cramer C, Truhlar D. Theoretical Calculation of Reduction Potentials. Organic Electrochemistry, Fifth Edition. 2015. pp. 229–259.

17. Baik M-H, Friesner RA. Computing Redox Potentials in Solution: Density Functional Theory as A Tool for Rational Design of Redox Agents. J Phys Chem A. 2002; 106: 7407–7412.

18. Park MS, Park I, Kang Y-S, Im D, Doo S-G. A search map for organic additives and solvents applicable in high-voltage rechargeable batteries. Phys Chem Chem Phys. 2016; 18: 26807–26815. https://doi.org/10.1039/c6cp05800k PMID: 27711632

19. Tagade PM, Adiga SP, Park MS, Pandian S, Hariharan KS, Kolake SM. Empirical Relationship between Chemical Structure and Redox Properties: Mathematical Expressions Connecting Structural Features to Energies of Frontier Orbitals and Redox Potentials for Organic Molecules. J Phys Chem C. 2018; 122: 11322–11333.

20. Llano J, Eriksson LA. First principles electrochemistry: Electrons and protons reacting as independent ions. J Chem Phys. 2002; 117: 10193–10206.

21. Isegawa M, Neese F, Pantazis DA. Ionization Energies and Aqueous Redox Potentials of Organic Molecules: Comparison of DFT, Correlated ab Initio Theory and Pair Natural Orbital Approaches. J Chem Theory Comput. 2016; 12: 2272–2284. https://doi.org/10.1021/acs.jctc.6b00252 PMID: 27065224

22. Winget P, Cramer CJ, Truhlar DG. Computation of equilibrium oxidation and reduction potentials for reversible and dissociative electron-transfer reactions in solution. Theor Chem Acc. 2004; 112. https://doi.org/10.1007/s00214-004-0577-0

23. Ree N, Andersen CL, Kilde MD, Hammerich O, Nielsen MB, Mikkelsen KV. The quest for determining one-electron redox potentials of azulene-1-carbonitriles by calculation. Phys Chem Chem Phys. 2018; 20: 7438–7446. https://doi.org/10.1039/c7cp08687c PMID: 29484319

24. Ho J. Are thermodynamic cycles necessary for continuum solvent calculation of pKas and reduction potentials? Phys Chem Chem Phys. 2015; 17: 2859–2868. https://doi.org/10.1039/c4cp04538f PMID: 25503399

25. Huskinson B, Marshak MP, Suh C, Er S, Gerhardt MR, Galvin CJ, et al. A metal-free organic-inorganic aqueous flow battery. Nature. 2014; 505: 195–198. https://doi.org/10.1038/nature12909 PMID: 24402280

26. Er S, Suh C, Marshak MP, Aspuru-Guzik A. Computational design of molecules for an all-quinone redox flow battery. Chem Sci. The Royal Society of Chemistry; 2015; 6: 885–893.

27. Gerhardt MR, Tong L, Gómez-Bombarelli R, Chen Q, Marshak MP, Galvin CJ, et al. Anthraquinone Derivatives in Aqueous Flow Batteries. Adv Energy Mater. 2017; 7. https://doi.org/10.1002/aenm.201700358

28. Pineda Flores SD, Martin-Noble GC, Phillips RL, Schrier J. Bio-Inspired Electroactive Organic Molecules for Aqueous Redox Flow Batteries. 1. Thiophenoquinones. J Phys Chem C. American Chemical Society; 2015; 119: 21800–21809.

29. Clayden J, Greeves N, Warren S, Wothers P. Organic Chemistry 2001. Oxford University Press;

30. Goldberg RN, Tewari YB, Bhat TN. Thermodynamics of enzyme-catalyzed reactions—a database for quantitative biochemistry. Bioinformatics. 2004; 20: 2874–2877. https://doi.org/10.1093/bioinformatics/bth314 PMID: 15145806

31. Alberty RA. Thermodynamics of Biochemical Reactions. John Wiley & Sons; 2005.

32. Schwabe T, Grimme S. Towards chemical accuracy for the thermodynamics of large molecules: new hybrid density functionals including non-local correlation effects. Phys Chem Chem Phys. 2006; 8: 4398–4401. https://doi.org/10.1039/b608478h PMID: 17001405

33. Grimme S. Semiempirical hybrid density functional with perturbative second-order correlation. J Chem Phys. 2006; 124: 034108. https://doi.org/10.1063/1.2148954 PMID: 16438568

34. Alberty RA, Cornish-Bowden A, Goldberg RN, Hammes GG, Tipton K, Westerhoff HV. Recommendations for terminology and databases for biochemical thermodynamics. Biophys Chem. 2011; 155: 89–103. https://doi.org/10.1016/j.bpc.2011.03.007 PMID: 21501921

35. Bar-Even A, Noor E, Flamholz A, Buescher JM, Milo R. Hydrophobicity and charge shape cellular metabolite concentrations. PLoS Comput Biol. 2011; 7: e1002166. https://doi.org/10.1371/journal.pcbi.1002166 PMID: 21998563

36. Noor E, Haraldsdóttir HS, Milo R, Fleming RMT. Consistent estimation of Gibbs energy using component contributions. PLoS Comput Biol. 2013; 9: e1003098. https://doi.org/10.1371/journal.pcbi.1003098 PMID: 23874165

37. Flamholz A, Noor E, Bar-Even A, Milo R. eQuilibrator—the biochemical thermodynamics calculator. Nucleic Acids Res. 2012; 40: D770–5. https://doi.org/10.1093/nar/gkr874 PMID: 22064852

38. Schneider N, Lowe DM, Sayle RA, Landrum GA. Development of a novel fingerprint for chemical reactions and its application to large-scale reaction classification and similarity. J Chem Inf Model. 2015; 55: 39–53. https://doi.org/10.1021/ci5006614 PMID: 25541888

39. Morgan HL. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. J Chem Doc. American Chemical Society; 1965; 5: 107–113.

40. Jean M, DeMoss RD. Indolelactate dehydrogenase from Clostridium sporogenes. Can J Microbiol. 1968; 14: 429–435. PMID: 4384683

41. Trinchant J-C, Rigaud J. Lactate Dehydrogenase from Rhizobium. Purification and Role in Indole Metabolism. Physiol Plant. Blackwell Publishing Ltd; 1974; 32: 394–399.

42. Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, et al. Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. Nat Chem Biol. 2011; 7: 445–452. https://doi.org/10.1038/nchembio.580 PMID: 21602812

43. Hadadi N, Hafner J, Shajkofci A, Zisaki A, Hatzimanikatis V. ATLAS of Biochemistry: A Repository of All Possible Biochemical Reactions for Synthetic Biology and Metabolic Engineering Studies. ACS Synth Biol. 2016; 5: 1155–1166. https://doi.org/10.1021/acssynbio.6b00054 PMID: 27404214

44. Bar-Even A, Flamholz A, Noor E, Milo R. Rethinking glycolysis: on the biochemical logic of metabolic pathways. Nat Chem Biol. 2012; 8: 509–517. https://doi.org/10.1038/nchembio.971 PMID: 22596202

45. Weber AL. Chemical constraints governing the origin of metabolism: the thermodynamic landscape of carbon group transformations under mild aqueous conditions. Orig Life Evol Biosph. 2002; 32: 333–357. PMID: 12458737

46. Cantor SM, Peniston QP. The Reduction of Aldoses at the Dropping Mercury Cathode: Estimation of the aldehydo Structure in Aqueous Solutions1. J Am Chem Soc. American Chemical Society; 1940; 62: 2113–2121.

47. Albe KR, Butler MH, Wright BE. Cellular concentrations of enzymes and their substrates. J Theor Biol. 1990; 143: 163–195. PMID: 2200929

48. Heineke D, Riens B, Grosse H, Hoferichter P, Peter U, Flügge UI, et al. Redox Transfer across the Inner Chloroplast Envelope Membrane. Plant Physiol. 1991; 95: 1131–1137. PMID: 16668101

49. Bekers KM, Heijnen JJ, van Gulik WM. Determination of the in vivo NAD:NADH ratio in Saccharomyces cerevisiae under anaerobic conditions, using alcohol dehydrogenase as sensor reaction. Yeast. 2015; 32: 541–557. https://doi.org/10.1002/yea.3078 PMID: 26059529

50. Zhao Y, Hu Q, Cheng F, Su N, Wang A, Zou Y, et al. SoNar, a Highly Responsive NAD+/NADH Sensor, Allows High-Throughput Metabolic Screening of Anti-tumor Agents. Cell Metab. 2015; 21: 777–789. https://doi.org/10.1016/j.cmet.2015.04.009 PMID: 25955212

51. Zhang J, ten Pierick A, van Rossum HM, Seifar RM, Ras C, Daran J-M, et al. Determination of the Cytosolic NADPH/NADP Ratio in Saccharomyces cerevisiae using Shikimate Dehydrogenase as Sensor Reaction. Sci Rep. 2015; 5: 12846. https://doi.org/10.1038/srep12846 PMID: 26243542

52. Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. Nat Chem Biol. 2009; 5: 593–599. https://doi.org/10.1038/nchembio.186 PMID: 19561621

53. Wächtershäuser G. Evolution of the first metabolic cycles. Proc Natl Acad Sci U S A. 1990; 87: 200–204. PMID: 2296579

54. Chen X, Alonso AP, Allen DK, Reed JL, Shachar-Hill Y. Synergy between (13)C-metabolic flux analysis and flux balance analysis for understanding metabolic adaptation to anaerobiosis in E. coli. Metab Eng. 2011; 13: 38–48. https://doi.org/10.1016/j.ymben.2010.11.004 PMID: 21129495

55. Fait A, Fromm H, Walter D, Galili G, Fernie AR. Highway or byway: the metabolic role of the GABA shunt in plants. Trends Plant Sci. 2008; 13: 14–19. https://doi.org/10.1016/j.tplants.2007.10.005 PMID: 18155636

56. Romano AH, Conway T. Evolution of carbohydrate metabolic pathways. Res Microbiol. 1996; 147: 448–455. PMID: 9084754

57. Siebers B, Schönheit P. Unusual pathways and enzymes of central carbohydrate metabolism in Archaea. Curr Opin Microbiol. 2005; 8: 695–705. https://doi.org/10.1016/j.mib.2005.10.014 PMID: 16256419

58. Lee HJ, Lee SH, Park CB, Won K. Coenzyme analogs: excellent substitutes (not poor imitations) for electrochemical regeneration. Chem Commun. 2011; 47: 12538–12540.

59. O'Brien PJ, Siraki AG, Shangari N. Aldehyde sources, metabolism, molecular toxicity mechanisms, and possible effects on human health. Crit Rev Toxicol. 2005; 35: 609–662. PMID: 16417045

60. Ferguson GP. Protective mechanisms against toxic electrophiles in Escherichia coli. Trends Microbiol. 1999; 7: 242–247. PMID: 10366861

61. Zimmermann M, Sauer U, Zamboni N. Quantification and mass isotopomer profiling of α-keto acids in central carbon metabolism. Anal Chem. 2014; 86: 3232–3237. https://doi.org/10.1021/ac500472c PMID: 24533614

62. Kuhn ML, Zemaitaitis B, Hu LI, Sahu A, Sorensen D, Minasov G, et al. Structural, kinetic and proteomic characterization of acetyl phosphate-dependent bacterial protein acetylation. PLoS One. 2014; 9: e94816. https://doi.org/10.1371/journal.pone.0094816 PMID: 24756028

63. Leventis R, Juel G, Knudsen JK, Silvius JR. Acyl-CoA binding proteins inhibit the nonenzymic S-acylation of cysteinyl-containing peptide sequences by long-chain acyl-CoAs. Biochemistry. 1997; 36: 5546–5553. https://doi.org/10.1021/bi963029h PMID: 9154938

64. Heller S, McNaught A, Stein S, Tchekhovskoi D, Pletnev I. InChI—the worldwide chemical structure identifier standard. J Cheminform. 2013; 5: 7. https://doi.org/10.1186/1758-2946-5-7 PMID: 23343401

65. Dosselaere F, Vanderleyden J. A metabolic node in action: chorismate-utilizing enzymes in microorganisms. Crit Rev Microbiol. 2001; 27: 75–131. https://doi.org/10.1080/20014091096710 PMID: 11450855

66. Teramoto H, Inui M, Yukawa H. Regulation of expression of genes involved in quinate and shikimate utilization in Corynebacterium glutamicum. Appl Environ Microbiol. 2009; 75: 3461–3468. https://doi.org/10.1128/AEM.00163-09 PMID: 19376919

67. Jinich A, Rappoport D, Dunn I, Sanchez-Lengeling B, Olivares-Amaya R, Noor E, et al. Quantum chemical approach to estimating the thermodynamics of metabolic reactions. Sci Rep. 2014; 4: 7022. https://doi.org/10.1038/srep07022 PMID: 25387603

68. Hadadi N, Ataman M, Hatzimanikatis V, Panayiotou C. Molecular thermodynamics of metabolism: quantum thermochemical calculations for key metabolites. Phys Chem Chem Phys. 2015; 17: 10438–10453. https://doi.org/10.1039/c4cp05825a PMID: 25799954

69. Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. MIT Press; 2006.

70. Dewar MJS, Trinajstic N. Ground states of conjugated molecules—XIV. Tetrahedron. 1969; 25: 4529–4534.

71. Johnsson Wass JRT, Tobias Johnsson J, Ahlberg E, Panas I, Schiffrin DJ. Quantum Chemical Modeling of the Reduction of Quinones. J Phys Chem A. 2006; 110: 2005–2020. https://doi.org/10.1021/jp055414z PMID: 16451036

72. Tazhigulov RN, Bravaya KB. Free Energies of Redox Half-Reactions from First-Principles Calculations. J Phys Chem Lett. 2016; 7: 2490–2495. https://doi.org/10.1021/acs.jpclett.6b00893 PMID: 27295124

73. Taylor M, Scott C, Grogan G. F420-dependent enzymes—potential for applications in biotechnology. Trends Biotechnol. 2013; 31: 63–64. https://doi.org/10.1016/j.tibtech.2012.09.003 PMID: 23098999

74. Eirich LD, Vogels GD, Wolfe RS. Distribution of coenzyme F420 and properties of its hydrolytic fragments. J Bacteriol. 1979; 140: 20–27. PMID: 40952

75. de Poorter LMI, Geerts WJ, Keltjens JT. Hydrogen concentrations in methane-forming cells probed by the ratios of reduced and oxidized coenzyme F420. Microbiology. 2005; 151: 1697–1705. https://doi.org/10.1099/mic.0.27679-0 PMID: 15870477

76. Zhang W, Zhang M, Gao C, Zhang Y, Ge Y, Guo S, et al. Coupling between d-3-phosphoglycerate dehydrogenase and d-2-hydroxyglutarate dehydrogenase drives bacterial l-serine synthesis. Proc Natl Acad Sci U S A. 2017; 114: E7574–E7582. https://doi.org/10.1073/pnas.1619034114 PMID: 28827360

77. Braakman R, Smith E. The emergence and early evolution of biological carbon-fixation. PLoS Comput Biol. 2012; 8: e1002455. https://doi.org/10.1371/journal.pcbi.1002455 PMID: 22536150

78. Besteiro S, Biran M, Biteau N, Coustou V, Baltz T, Canioni P, et al. Succinate secreted by Trypanosoma brucei is produced by a novel and unique glycosomal enzyme, NADH-dependent fumarate reductase. J Biol Chem. 2002; 277: 38001–38012. https://doi.org/10.1074/jbc.M201759200 PMID: 12138089

79. Miura A, Kameya M, Arai H, Ishii M, Igarashi Y. A soluble NADH-dependent fumarate reductase in the reductive tricarboxylic acid cycle of Hydrogenobacter thermophilus TK-6. J Bacteriol. 2008; 190: 7170–7177. https://doi.org/10.1128/JB.00747-08 PMID: 18757546

80. Bar-Even A, Noor E, Lewis NE, Milo R. Design and analysis of synthetic carbon fixation pathways. Proc Natl Acad Sci U S A. 2010; 107: 8889–8894. https://doi.org/10.1073/pnas.0907176107 PMID: 20410460

81. Schwander T, Schada von Borzyskowski L, Burgener S, Cortina NS, Erb TJ. A synthetic pathway for the fixation of carbon dioxide in vitro. Science. 2016; 354: 900–904. https://doi.org/10.1126/science.aah5237 PMID: 27856910

82. Haraldsdóttir HS, Thiele I, Fleming RMT. Quantitative assignment of reaction directionality in a multi-compartment human metabolic reconstruction. Biophys J. 2012; 102: 1703–1711. https://doi.org/10.1016/j.bpj.2012.02.032 PMID: 22768925

83. Jankowski MD, Henry CS, Broadbelt LJ, Hatzimanikatis V. Group contribution method for thermodynamic analysis of complex metabolic networks. Biophys J. 2008; 95: 1487–1499. https://doi.org/10.1529/biophysj.107.124784 PMID: 18645197

84. Henry CS, Broadbelt LJ, Hatzimanikatis V. Thermodynamics-based metabolic flux analysis. Biophys J. 2007; 92: 1792–1805. https://doi.org/10.1529/biophysj.106.093138 PMID: 17172310

85. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, et al. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol. EMBO Press; 2007; 3: 121.

86. Neese F. The ORCA program system. WIREs Comput Mol Sci. John Wiley & Sons, Inc.; 2012; 2: 73–78.

87. Klamt A, Schüürmann G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J Chem Soc Perkin Trans 2. The Royal Society of Chemistry; 1993; 0: 799–805.

88. Grimme S, Antony J, Ehrlich S, Krieg H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J Chem Phys. 2010; 132: 154104. https://doi.org/10.1063/1.3382344 PMID: 20423165

89. Becke AD, Johnson ER. Exchange-hole dipole moment and the dispersion interaction revisited. J Chem Phys. 2007; 127: 154108. https://doi.org/10.1063/1.2795701 PMID: 17949133

90. Goerigk L, Grimme S. Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals-Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Non-covalent Interactions. J Chem Theory Comput. 2011; 7: 291–309. https://doi.org/10.1021/ct100466k PMID: 26596152

91. Roch LM, Baldridge KK. Dispersion-Corrected Spin-Component-Scaled Double-Hybrid Density Functional Theory: Implementation and Performance for Non-covalent Interactions. J Chem Theory Comput. 2017; 13: 2650–2666. https://doi.org/10.1021/acs.jctc.7b00220 PMID: 28537392

92. Rogers D, Hahn M. Extended-connectivity fingerprints. J Chem Inf Model. 2010; 50: 742–754. https://doi.org/10.1021/ci100050t PMID: 20426451