Chemiosmosis and Microscopic Reversibility: Irreconcilable. Thermodynamic limitations of coupled equilibria and the role of ATP

Sosale Chandrasekhar*

Department of Organic Chemistry, Indian Institute of Science, Bangalore 560 012 (India)

*E-mail: sosale@orgchem.iisc.ernet.in; sosalechandra@hotmail.com

ABSTRACT

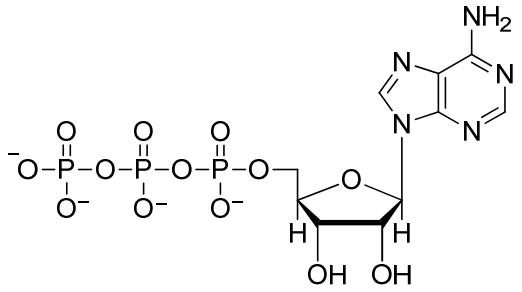
It is argued that the chemiosmotic theory, in positing the egress and ingress of protons across a membrane via different routes, violates the principle of microscopic reversibility (PMR). The theory was proposed as a mechanism for coupling the exergonic oxidation of NADH to the endergonic synthesis of ATP, believed to be a ready-to-use form of biochemical energy. The key proton-pump mechanism, however, should also operate in reverse by the PMR, hence is unlikely to be valid. Generally, the transfer of free energy *via* coupled equilibria is problematical, as neither of the coupled processes would go to completion. (However, it may be feasible in the case of entropy-driven reactions, as happens in the reactions of the Krebs cycle that lead to the formation of NADH.) Furthermore, reported thermodynamic data do not support the idea of ATP as a 'high energy' molecule, but rather indicate that ATP hydrolysis is driven by release of phosphate. In fact, many of the reactions believed to be thermodynamically driven by ATP hydrolysis are likely exergonic *per se*. ATP possibly provides part of the activation energy for the reaction in many cases. An alternative (non-equilibrium) mechanism for ATP synthesis is proposed based on the increase in pH that would accompany the oxidation of NADH. This would generate ADP^{3-} and HPO_4^{2-} which could form ATP^{4-} in a thermodynamically favoured process within hydrophobic pockets.

INTRODUCTION

The ingestion of foodstuffs, their oxidation and the storage of the resulting energy in a retrievable form constitute a defining characteristic of the higher life-forms. This cluster of activities enables other characteristics that define life, primarily locomotion, reproduction and neural function. Whilst fat is considered to represent the 'heavy-duty' mode of long term energy storage, it is not suitable as a quick-and-ready medium of energy exchange. Currently, the function of an easy and convenient energy currency in biological systems is widely believed to be performed by adenosine triphosphate (ATP, **1,** Scheme 1).

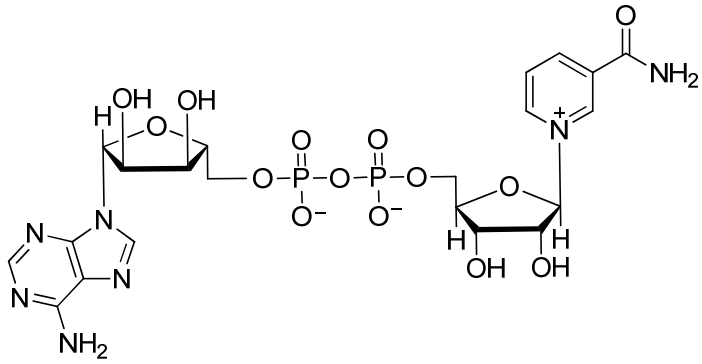
More often than not, however, ATP production is not directly linked to the oxidation of foodstuffs. All catabolic pathways apparently converge on to the tricarboxylic acid cycle, which performs the final oxidation of an acetyl unit to CO_2 and water. (The tricarboxylic acid cycle is also termed 'TCA cycle', 'citric acid cycle' and 'Krebs cycle', and is believed to operate in the mitochondrion.) A cyclical mechanism for oxidative degradation enables the regeneration of 'carrier' intermediates without the reversal of the individual steps, thus allowing the energy generated in a particular step to exit the system. A cyclical reversal is enabled by the continual input of the high energy acetyl units into the open system. (This 'reboots' oxaloacetate to citrate: details of the Krebs cycle may be found in standard textbooks of biochemistry, and need not be repeated here.)

The energy exiting the Krebs cycle is initially trapped by converting a molecule of nicotinamide adenine dinucleotide (NAD^+, **2a**) to its reduced form (NADH, **2b**). (In one of the steps a flavin analog is reduced.) For various reasons, however, NADH is unsuited to the role of a ubiquitous energy currency; hence it is re-oxidized in order to drive the formation of ATP, which is disbursed to the tissues and organs constituting living systems. (Essentially, NADH lacks the kinetic stability and modes of reactivity possessed by ATP.)



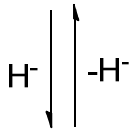
ATP

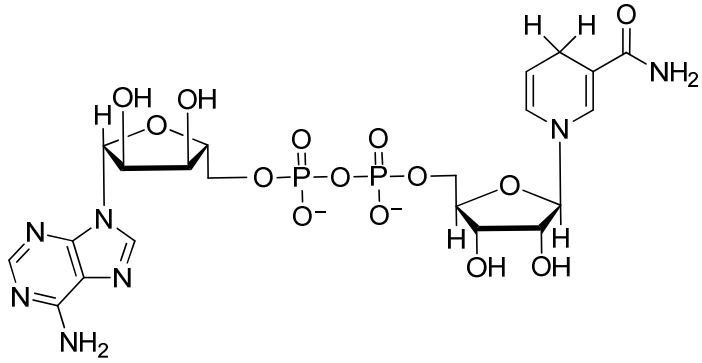
1



NAD^+

2a





NADH

2b

Scheme 1. *The structures of adenosine triphosphate (**1**) and nicotinamide adenine dinucleotide (**2a** and **2b,** oxidized and reduced forms respectively), key molecules mediating metabolic energy storage and transfer*

An alternative catabolic mode, in which ATP is directly produced during oxidation, is found in glycolysis. Under aerobic conditions, this involves the oxidation of glucose to pyruvate, by a multistep non-cyclic mechanism. In two of the oxidative steps ATP is directly produced by a process termed 'substrate-level phosphorylation', distinguishing it from the 'oxidative phosphorylation' occurring in the Krebs cycle. Although NADH is produced in one of the steps in glycolysis, it does not leave the system but is re-oxidized to NAD^+, hence works in a catalytic sense. (It is noteworthy that glycolysis represents an early stage in the catabolism of the complex glucose molecule, so a cyclic mechanism is not feasible. In contrast, the Krebs cycle involves the annihilation of a simple acetate unit! Again, details may be found in standard texts.)

The mechanism by which the energy released in the oxidation of NADH is coupled to the formation of ATP (*vide supra*) has been a prime concern of physical biochemistry for several decades. Currently, it is believed that this is accomplished *via* a chemiosmotic process, by which the energy released from NADH is used to eject protons across a membrane, thus creating an imbalance in the proton concentrations on either side of the membrane. This translates into a potential energy difference across the membrane. This can be relieved if the protons re-cross the membrane, but at a different site at which the released energy can drive ATP production.

Apparently, analogous ideas had been employed earlier in the case of electrochemical cells, although their extension to biological systems is interesting. All the same, there are several problems with the chemiosmotic hypothesis, both in the gross and in the detail. A particular problem is that the mechanism involves a 'physical traverse', in the sense that the protons are conducted along a specified path across the membrane. This raises questions about energy transfer and its reversibility in the context of the principle of microscopic

reversibility (PMR). These derive from more general questions pertaining to the coupling of exergonic and endergonic reactions, so frequently employed in biological processes.

In fact, these problems derive from the nature of the ATP molecule itself. For long, it has been considered as a 'high energy' molecule because of the presence of the triphosphate unit. However, there is little evidence to support this view, or even that ATP is employed for driving thermodynamically disfavoured reactions. So, how is ATP synthesised *in vivo* and what is its role therein? The discussion below attempts to answer this clearly weighty question in the light of available evidence.

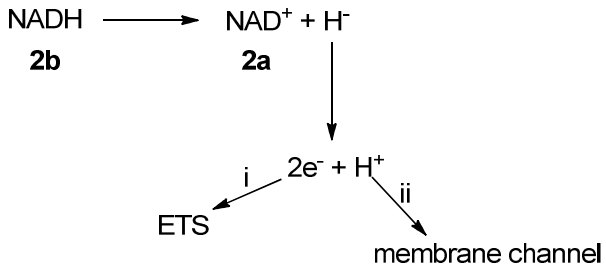
DISCUSSION

The oxidation of NADH and the electron transport system (ETS)

The overall oxidation of NADH (**2b**), in the context of the Krebs cycle, may be represented by Eq. 1. Note that the reaction is usually represented by including a proton, so the by-product is H_2O rather than OH^-: however, the equation would then be unbalanced in terms of charge, as the source of the H^+ is unclear. Eq. 1 below appears a more accurate representation of the reaction, and also implies that the oxidation involves an increase in pH, at least at the reaction site. (The mode of neutralization of the OH^- has an important bearing on a possible mechanism for the synthesis of ATP, as will be discussed further below.)

Although the mechanism pertaining to Eq. 1 is of intriguing complexity, it may be represented by the formal release of a hydride ion from the dihydropyridine ring in **2b,** followed by the release of an electron pair from the hydride ion to an electron-transport system (ETS). This process produces the pyridinium ion moiety residing in **2a,** apart from a proton and the electron pair captured by the ETS (Scheme 2 and Fig. 1).

$$NADH + 0.5O_2 \rightarrow NAD^+ + OH^- \qquad (1)$$



Scheme 2. *Formal representation of the oxidation of NADH (**2b**) via the formation of NAD^+ (**2a**) and a putative hydride ion (H^-); this is stripped of its electron pair which is conducted via the electron transport system (ETS, path i), the proton by-product being pumped out of the mitochondrial membrane (path ii, cf. Fig. 1)*

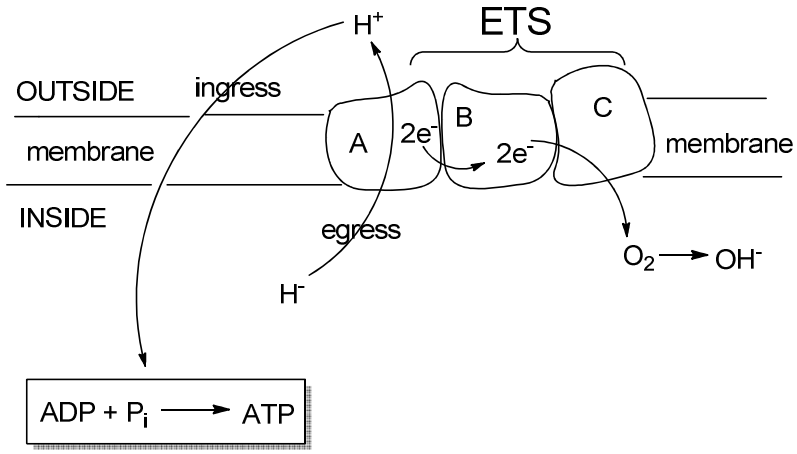


Fig. 1. *The key elements of the chemiosmotic mechanism (cf. Scheme 2) involving the mitochondrial membrane and some of the carriers (A, B and C) of the electron transport system (ETS); a proton exits via one channel and re-enters via another, possibly violating the principle of microscopic reversibility*

The ETS is believed to be constituted of a series of electron acceptors of increasing reduction potential, which thus enables the conduction of the initially captured electron pair down to the last acceptor, *i.e.* the O_2 molecule. All the acceptors except the last (O_2), also function as donors to the next-in-line acceptor. It is particularly noteworthy that the initial stages of the conduction are driven by the relatively low reduction potentials of the donors,

whereas the final stages are driven by the relatively high reduction potentials of the acceptors. This makes all the steps highly exergonic and the transfer of the electron pair irreversible at all stages. The highly organized nature of the ETS, in which the electron acceptor-donor units are embedded in a protein matrix, prevents diffusion, so the electrons are constrained to be conducted as above.

However, central to the chemiosmotic mechanism is the fate of the proton that is by-produced in the above process. It is currently believed that the exergonic conduction of the electron pair along the ETS drives the initial stripping of the electrons from the putative hydride ion and, furthermore, the channelling of the resulting proton across a surrounding membrane. The expelled proton thus represents a high energy species poised to perform useful work – the synthesis of ATP.

For this to be accomplished, the proton is smuggled back into the original side of the membrane, but *via* a different entry point. At the interior terminus of this channel, is located the molecular apparatus for the synthesis of ATP (**1**) from adenosine diphosphate (ADP, the diphosphate analog of **1**) and inorganic phosphate (P_i). Thus, the energy from the exergonic oxidation of NADH is finally consummated in the endergonic synthesis of ATP. This, of course, also re-establishes the original osmotic equilibrium, in terms of the proton concentrations, across the membrane. Thus are the broad contours of the chemiosmotic theory for the production of ATP, the purported high energy metabolic energy storehouse.

A rather disturbing feature of the above mechanism, however, is the fact that the proton is expelled across the membrane *via* one channel but retrieved to the interior *via* another. This possibly violates the fundamental principle of microscopic reversibility, which must be obeyed in all chemical reactions, as discussed below.

The principle of microscopic reversibility (PMR)

The principle of microscopic reversibility (PMR) essentially ensures the uniqueness of the equilibrium constant of a reversible reaction, under a defined set of conditions. Originally proposed as the principle of detailed balance, the PMR has many equivalent formulations. All of these essentially imply that any chemical event, such as a molecular collision, bond-cleavage or bond-making, etc., occurs in both 'forward' and 'reverse' directions, and does so with equal probability at equilibrium.

There are two ways of applying the PMR in the context of the above chemiosmotic theory of ATP production. One way would be to demand that the path followed by the proton to the outside of the membrane be the same as the path for its re-entry. Alternatively, it may be demanded that the entire traverse across both the trans-membrane channels be reversible. Intriguingly, not only would both demands accord with the PMR, but also, they would either not lead to the synthesis of ATP or reverse it!

This, of course, begs the question: would they reverse the oxidation of NADH? However, this is not as valid a riposte as it may sound, bearing in mind that the synthesis of ATP is about as endergonic as the oxidation of NADH is exergonic! In other words, the intuitive feeling that the oxidation of NADH is irreversible must be balanced against the cold logic that ATP is about equally thermodynamically unstable! On these grounds, therefore, both processes can be reversible, and indeed must be so by the PMR.

A fine detail and useful insights

An interesting fine detail of the chemiosmotic theory is provided by the so-called Q cycle. This derives its name from the redox coenzyme Q that mediates the critical electron transfer step between the iron-sulphur protein complex and the series of cytochromes in the ETS. In light of the above arguments, all the reactions comprising the Q cycle would be reversible by the PMR.

Also, an interesting feature of the chemiosmotic mechanism (Scheme 2) is the presence of a branching point, beyond which the reaction is bifurcated into the electron transfer (path i) and proton extrusion (path ii) routes. Whilst path i is exergonic, path ii is endergonic, the chemiosmotic theory proposing that the latter is assisted by the former. This would require, of course, some mechanism for the transfer of energy between the two paths. By the PMR, however, this transfer of energy would be reversible.

Thus, the initial stages of the electron transfer route are reversible by the PMR. Indeed, this reversal can now be assisted by the reversal of the proton extrusion, which would now be exergonic! Also, it would appear that such assistance as above would be merely kinetic rather than thermodynamic, as this would require the stabilization of the by-product (*i.e.* the extruded proton).

More generally, an exergonic reaction would be propelled in a direction that is thermodynamically 'downhill'. However, if it is coupled to an endergonic reaction, the original exergonic reaction could well be propelled in the opposite direction.

Intriguingly, the thermodynamic coupling of reactions is rarely considered *in vitro*. However, the idea is pervasive in the case of biochemical processes, as such coupling is believed to drive reactions 'uphill' at constant temperature. However, the physico-chemical basis of such thermodynamically-coupled processes is intriguing, as discussed below.

Thermodynamic consequences of coupled equilibria

The coupling of two reversible equilibria, involving the species A, B and C, is shown in Eqs. 2 and 3 (Scheme 3). If K_1 and K_2 are the original equilibrium constants as shown, the new equilibrium constant for the overall coupled equilibrium (Eq. 4) would be $K = (K_1K_2)$. Likewise, if the Gibbs free energy changes are ΔG^{o}_1 and ΔG^{o}_2 (corresponding to K_1 and K_2 respectively), the overall Gibbs free energy change corresponding to K would be $\Delta G^{o} = (\Delta G^{o}_1 + \Delta G^{o}_2)$.

$$A \rightleftharpoons B \qquad (2)$$

$$B \rightleftharpoons C \qquad (3)$$

$$A \rightleftharpoons C \qquad (4)$$

$K_1 = [B]/[A]$; $K_2 = [C]/[B]$; $K = K_1K_2 = [C]/[A]$

Scheme 3. *The coupling of the two equilibria (Eqs. 2 and 3) leading to the overall process in Eq. 4; the K's are the relevant equilibrium constants as shown*

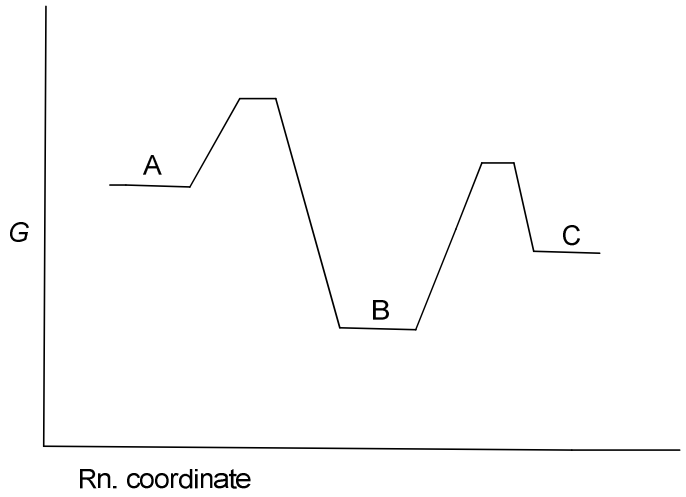


Fig. 2. *The energy profile for the coupled reactions shown in Scheme 3, when the conversion of A to B is exergonic and that of B to C is endergonic; the ratio A:B:C is governed by their relative free energy contents*

Now, if the conversion of A to B is exergonic ($\Delta G^o_1 < 0$) and that of B to C is endergonic ($\Delta G^o_2 > 0$), the sign and magnitude of ΔG^o would be determined by the relative values of ΔG^o_1 and ΔG^o_2. If for instance, $\Delta G^o_1 \sim -\Delta G^o_2$, $\Delta G^o \sim 0$ and $K \sim 1$, so the conversion of A to C would not go to completion. Also, there would be a substantial build up of the intermediate B, as this is the most stable species in the mixture. These arguments are depicted in the energy profile diagram in Fig. 2.

It is also noteworthy, in general, that an exergonic reaction is essentially driven to completion by an overall increase in the entropy of the system and its surroundings. Thus, the enthalpy released to the surroundings in an isothermal process is manifested as a corresponding increase in its entropy. This, however, is obviated if the exergonic process is coupled to another endergonic process. In such a case, it seems unlikely that either process would go to completion.

Also, endergonic reactions cannot be driven to completion by thermal energy alone, as this will not entirely circumvent the unfavourable equilibrium constant. Interestingly, neither can endergonic reactions be driven to completion by the transfer of chemical potential energy – accruing from the Gibbs free energy released by another exergonic reaction – as indicated by the arguments presented above. (Endergonic reactions are normally driven to completion *in vitro* by recourse to le Chatelier's principle, *i.e.* employing an excess reactant or removal of a product.)

These arguments ultimately lead to the PMR, by which all such energy transfers must be reversible. Thus, the final equilibrium state is defined by the relative Gibbs free energy contents of all the participating species, as also the total energy within the system. Intriguingly, examples of the coupling of exergonic and endergonic reactions *in vitro* appear to be almost non-existent. In fact, the above discussed case (*cf.* Scheme 3 and Fig. 2) is not an accurate representation of such coupling, when the transfer of Gibbs free energy between different reactions is involved, *e.g.* ATP synthesis from the oxidation of NADH.

A mechanical analogy (not shown), however, is useful in clarifying the arguments. This involves a piston-cylinder arrangement (A_1) containing a gas at a specified pressure (P_1) that is externally connected to another piston-cylinder arrangement (A_2) with a gas at pressure (P_2). The two pistons are positioned to act against each other, so mechanical work may be performed.

If $P_1 \sim P_2$, little work will be produced, with minimal movements of the pistons. If $P_1 > P_2$, A_1 will perform work on A_2 until the two pressures equalize. It is only when $P_1 >>> P_2$, that A_1 would undergo nearly complete expansion with maximal movement of the piston. This analogy corresponds to the coupling of an overwhelmingly exergonic process to a marginally endergonic one (*vide supra*).

Coupled equilibria in biochemical processes: ATP synthesis

The above arguments lead to interesting if intriguing results when applied to biochemical reactions. The trapping of the energy released in the oxidation of NADH by the synthesis of ATP is a cornerstone of modern chemical biology. Although the precise mechanism of energy transfer remains unclear, it is believed that large quantities of ATP – equal to a person's body weight daily, by one estimate – are produced thus. It is interesting to apply the general principles derived above to this purported phenomenon.

Thus, if the Gibbs free energy change in the oxidation of NADH is similar in magnitude to the Gibbs free energy change in the synthesis of ATP, neither of the reactions would go to completion. This can be seen by writing down a combined equation for the two reactions, involving all the reactant species, as in Eq. 6. This is obtained by adding the two reactions composing the overall process, the oxidation of NADH (Eq. 1) and the synthesis of ATP from ADP and P_i (Eq. 5).

$$NADH + 0.5O_2 \rightarrow NAD^+ + OH^- \qquad (1)$$

$$ADP + P_i \rightarrow ATP + H_2O \qquad (5)$$

$$NADH + 0.5O_2 + ADP + P_i \rightarrow NAD^+ + OH^- + ATP + H_2O \qquad (6)$$

Although Eq. 6 is somewhat hypothetical, as all the species shown on the left hand side do not necessarily react with each other, it is a valid formal representation of the overall changes, particularly involving the Gibbs free energy.

The above arguments then indicate that the overall process represented in Eq. 6 would not go to completion, if the Gibbs free energy changes in Eqs. 1 and 5 are similar in magnitude. This, in fact, manifests as a violation of the PMR in the chemiosmotic mechanism, which – apparently – assumes that the process would go to completion.

However, the process in Eq. 6 can be driven to completion if the reaction in Eq. 1 were to be very highly exergonic, so the overall process in Eq. 6 would also be highly exergonic. (This would derive from a relatively high Gibbs free energy content of NADH.) In this case, however, a considerable amount of the energy released in the oxidation of NADH (Eq. 1) would be wasted to the surrounding environment. (This would be equal to the free energy difference between A and C, *cf.* Fig. 2.)

This, of course, raises fundamental questions not only about the efficiency of the process, but also as to whether evolution would have taken such an inefficient course. (Thus, a violation of the PMR could be avoided at great energetic cost!)

How exergonic is the oxidation of NADH?

Apparently, the key to resolving the above controversy would lie in determining how exergonic the oxidation of NADH really is. Intriguingly, there appears to be no definitive answer to this question. The oxidation of NADH is essentially driven by the aromaticity of the pyridinium moiety in the resulting NAD^+ and the reduction of oxygen (Eq. 1). However, thermodynamic data, particularly as applies to *in vivo* conditions, is apparently inconclusive or scarce.

It is currently believed that the oxidation of one molecule of NADH leads to the formation of 3 molecules of ATP (from ADP and P_i). Since the hydrolysis of ATP is believed to be exergonic by ~30 kJ mol^{-1} under physiological conditions, the oxidation of NADH must be exergonic by at least ~90 kJ mol^{-1}. Also, each molecule of ATP is apparently synthesized at different stages of the conduction of the electrons through the ETS. This implies that each

of the steps is only marginally exergonic – relative to the overall process – including the final step involving O_2. Most importantly, this also implies that the overall process may, in principle, be reversed in stages by appropriate input of energy.

Furthermore, the exergonicity of the oxidation of NADH has been estimated from the electrochemical reduction potentials of the NAD^+/NADH and O_2/H_2O cells. These are -0.32 V and 0.8166 V respectively. This leads to a combined voltage of 1.14 V (for the cells coupled back-to-back), and a corresponding free energy change of -220 kJ mol^{-1}. However, the applicability of these data to physiological conditions *in vivo* is unclear. Also, on the basis that the oxidation of a molecule of NADH leads to the synthesis of 3 molecules of ATP (~ 90 kJ mol^{-1} in sum, *vide supra*), their coupling would lead to a wastage of > 100 kJ mol^{-1}. (The above data relate to a standard state of 1 *M* and corrections for concentration changes need to be included, but they do provide reasonable estimates.)

This, of course, would drive the overall coupling process to completion, thus also avoiding a violation of the PMR: however, at an enormous cost in terms of energy! Again, would Nature be so wasteful?

Coupled reactions in the Krebs cycle: fundamentally revealing!

In fact, the Krebs cycle also employs coupled reactions for the synthesis of 3 NADH molecules. These (and an FADH molecule) are by-products of oxidative degradation steps which reduce NAD^+, thus representing another example of the capture of free energy. In light of the above critique of such coupling strategies, the question arises as to how thermodynamically efficient the synthesis of NADH itself is in the Krebs cycle.

Interestingly, two of the above synthetic steps are oxidative decarboxylations. These are clearly driven by the increase in entropy afforded by the release of gaseous CO_2, hence would be highly exergonic. In fact, such entropy-driven processes offer immense advantages, essentially because they do not release energy to the surroundings; also, they allow for the

formation of highly reactive by-products. Fascinatingly, both these attributes are apparently critical to the successful coupling of exergonic and endergonic processes.

As was argued at length above, coupling an exergonic reaction to an endergonic one is a useful strategy only if the former is very highly exergonic, so the overall coupled process would also be considerably exergonic. In such a case, however, the overall coupled process would go to completion, but at great energetic cost. It appears that in the case of the Krebs cycle, Nature has devised a strategy to circumvent this thermodynamic limitation by involving a large entropic component in the free energy expended to the surroundings. (The interpretation of entropy in such cases can be contentious: a fuller discussion can be found in the Appendix.)

ATP: powerhouse or paradox?

However it is synthesized, the view that ATP is a high-energy molecule pervades modern biochemical theory. There are at least two problems with this inexact viewpoint. Firstly, whether the energy content of a molecule is high or low is clearly a relative question. Secondly, even the absolute energy content of a molecule, *i.e.* in terms of its thermochemical heat of formation, often is no indicator of its reactivity and its preferred mode of decomposition under given conditions. In the case of ATP, its hydrolysis (Eq. 7) is exergonic under physiological conditions. This has led to the idea that the analogous decomposition of ATP is employed to drive endergonic reactions.

$$ATP + H_2O \rightarrow ADP + P_i \qquad (7)$$

As was seen above, there are fundamental problems involved in the thermodynamic coupling of exergonic and endergonic processes. Primarily, such coupled processes apparently may not go to completion. Furthermore, and intriguingly, it appears that many of the processes that ATP purportedly drives ‘uphill’ may well be inherently exergonic!

A case in point is protein synthesis, a process in which ATP has been implicated. Proteins, however, are composed of a large number of peptide units, which are known to be highly thermodynamically stable by virtue of the amide bond. (Entropy losses attending the condensation of the constituent amino acids would be minimized by the release of a water molecule in each step.) Further stabilization is also afforded by folding and aggregation, leading to secondary, tertiary and quaternary structures. Similar arguments apply to the case of oligonucleotides, which often rival the proteins in stability particularly in base. (The DNA duplex structure is extremely stable at normal temperatures.)

In fact, it may well be argued that these macromolecules must possess considerable thermodynamic stability in order to perform their appointed and critical biological functions! Otherwise, their formation would easily be reversed in the presence of a large excess of water *in vivo*, leading to their rapid breakdown!

Interestingly, the implication of ATP in the *in vivo* synthesis of macromolecules may indicate that ATP provides part of the activation energy for their formation. This is seen in the activation of the carboxyl and phosphoryl groups during oligopeptide and oligonucleotide syntheses respectively. The resulting mixed anhydride intermediates are rapidly formed, largely because of the reactivity of ATP, their further reactions also being rapid. The overall exergonic hydrolysis of ATP, apparently, adds to the exergonicity of the coupling process: however, this is not critical to the formation of the products, which are stable in themselves.

The above additional exergonicity indeed appears wasteful, as biological evolution is highly conservative in terms of energy usage! An intriguing possibility, however, is that the above macromolecular biosyntheses occur in relatively hydrophobic regions, in which the hydration of the inorganic phosphate by-product (derived from ATP) is minimal. As this hydration likely contributes significantly to the exergonicity of ATP hydrolysis (*vide infra*), the hydrolysis may be only marginally exergonic in a hydrophobic matrix.

Therefore, in the possibly hydrophobic environment in which the above biosyntheses occur, ATP is likely to provide kinetic – rather than any significant thermodynamic – support for the processes. (Intriguingly, this implies that ATP acts in a catalytic sense in these reactions: it would be worthwhile indeed to look for possible supporting evidence.)

How, then, is ATP made?

The above thermodynamic unravelling of ATP hydrolysis also leads to interesting insights into the likely mode of ATP synthesis. To recapitulate (*vide supra*), the problem is one of how to transfer the free energy from the exergonic oxidation of NADH to the endergonic synthesis of ATP, with minimal wastage to the surroundings. Although the above discussion is apparently bewildering as it flies in the face of currently cherished views, reported thermodynamic data offer interesting clues to address the problem. (Relevant data have been compiled in Table 1.)[1]

$$ADP^{3-} + HPO_4^{2-} + H^+ \rightarrow ATP^{4-} + H_2O \qquad (8)$$

To begin with, although the hydrolysis of ATP is exergonic under physiological conditions, it is not always so! Thus, under standard conditions, ATP synthesis can even be marginally exergonic by ~3 kJ mol^{-1} (Eq. 8). This refers to conditions of zero ionic strength and unspecified pH.

Also, the normally exergonic hydrolysis of ATP is apparently driven by the stability of the inorganic phosphate by-product, as – perhaps intriguingly – ADP is thermodynamically less stable than ATP by > 850 kJ mol^{-1}. In fact, it appears that the phosphate moiety confers substantial thermodynamic stability upon being appended to any molecule, *e.g.* glucose 6-phosphate is more stable than glucose by ~ 900 kJ mol^{-1}.

---

1. Alberty, R. A.; Goldberg, R. N. *Biochemistry* **1992,** *31*, 10610-10615.

Table 1.[1] The standard Gibbs free energy of formation ($\Delta_f G^o$) for relevant compounds (in kJ mol^{-1} and at temperature 298.15 K)[#]

Entry	Compound	$\Delta_f G^o$	
		a	**b**
1	Adenosine triphosphate (ATP^{4-})	-2573	-2097
2	Adenosine triphosphate ($HATP^{3-}$)	-2616	-2094
3	Adenosine diphosphate (ADP^{3-})	-1711	-1229
4	Adenosine diphosphate ($HADP^{2-}$)	-1752	-1225
5	Phosphate ($H_2PO_4^-$)	-1137	-1056
6	Phosphate (HPO_4^{2-})	-1096	-1058
7	H_2O	-237.0	-156.0
8	H^+	0.00	-0.81

[#](**a**: ionic strength = 0, pH unspecified; **b:** ionic strength = 0.25 *M*, pH = 7)

Generally, the relevant thermodynamic stability order appears to be: ATP > ADP > AMP. Thus, the hydrolytic reactivity of these species is apparently driven by substantial contributions by phosphate and by water. (The latter apparently manifests the greatest variation with ionic strength and pH, as may be expected.)

Interestingly, the fact that the reaction in Eq. 8 is marginally exergonic indicates the possibility that ATP synthesis can occur under the specified conditions. These, in fact, appear to approximate those to be found in a hydrophobic environment. It is interesting to note that, under similar conditions, the analogous reaction in Eq. 9 is endergonic by ~35 kJ mol^{-1}.

$$HADP^{2-} + H_2PO_4^- \rightarrow HATP^{3-} + H_2O \qquad (9)$$

This reaction involves the less ionized congeners of the phosphoryl species in Eq. 8. Under physiological conditions, however, the less ionized and more ionized congeners of each species would be present in nearly equal amounts. This is based on the relevant pK_a

values for ATP (~6.5), ADP (~6.5) and H_3PO_4 (7.2, 12.3). Therefore, the reaction in Eq. 8 is feasible only at pH values well above physiological levels (~7.0), when all the species shown would be present in substantial quantities.

In other words, it appears that ATP synthesis is only feasible at a relatively high pH. Intriguingly, this accords very well with the fact that the oxidation of NADH is accompanied by an increase in pH (Eq. 1)! Clearly, the deprotonation of the various species shown in Eq. 9 by the hydroxide ion produced in the oxidation of NADH (Eq.1), would not only buffer the medium (thus keeping the pH close to physiological levels), but also drive ATP synthesis by the reaction in Eq. 8.

The reaction in Eq. 8 is, of course, only marginally exergonic, which implies an equilibrium constant close to unity. All the same, it could lead to significant levels of ATP production. The reaction also allows for the possibility of selectively removing ATP from the reaction site, thus driving the equilibrium towards ATP formation. It is noteworthy that the reaction in Eq. 8 is enabled, relative to that in Eq. 9, by the higher reactivity of both ADP^{3-} and HPO_4^{2-}. (The products in Eq. 9 are, in fact, thermodynamically more stable overall than those in Eq. 8.)

Thus, the formation of the more highly ionized reactant species is critical to the success of the reaction in Eq. 8, which indicates the importance of a relatively high pH at the reaction site. Once formed, ATP possesses substantial kinetic stability so can diffuse away from the reaction site, and persist until kinetically activated by the appropriate enzyme.

The overall reaction for the formation of ATP from $HADP^{2-}$ and $H_2PO_4^-$ at high pH is shown in Eq. 10, along with the equilibrium constant K_3.

$$HADP^{2-} + H_2PO_4^- + OH^- \rightarrow ATP^{4-} + 2H_2O \qquad (10)$$

$$K_3 = [ATP^{4-}]/\{[\ HADP^{2-}][H_2PO_4^-][OH^-]\}$$

The fact that the oxidation of a molecule of NADH leads to the formation of 3 molecules of ATP, apparently indicates that K_3 ~3, under the extant physiological conditions. This can be understood by noting that K_3 possesses dimensions of M^{-2}: with an abundant supply of $HADP^{2-}$ and $H_2PO_4^-$, OH^- would be all but consumed so $[OH^-]$ would be minimal. Therefore, these conditions would establish the equivalence between a molecule of NADH oxidized and the number of ATP molecules produced.

ATP synthesis can only occur under non-equilibrium conditions!

It is noteworthy that ATP synthesis can only occur under non-equilibrium conditions, as ATP is thermodynamically unstable under normal physiological conditions. The mechanism proposed herein meets this essential criterion in invoking the possibility that ATP synthesis can occur in isolated pockets, away from the bulk medium that defines the equilibrium condition. (There is an inversion of the equilibrium constant for ATP synthesis within these pockets, *vide supra* and *cf.* Eqs. 8 and 9.) The possibility of the hydrolysis of ATP, upon its diffusion into the bulk medium, is countervailed by its kinetic stability, thus extending the non-equilibrium condition.

In contrast, the chemiosmotic mechanism essentially invokes an equilibrium based model, in which the exergonic oxidation of NADH is coupled to the endergonic synthesis of ATP (Fig. 1). This is thermodynamically implausible, as the reverse process is also possible: this is manifested as a violation of the PMR. Thus, by the PMR, the hydrolysis of ATP should reverse the flow of protons to the membrane exterior, thence towards the electron transport region itself. This should slow down and eventually staunch the electron transfer process that defines the initial stages of the oxidation of NADH.

CONCLUSIONS

The chemiosmotic mechanism for the coupling of the exergonic oxidation of NADH to the endergonic synthesis of ATP, appears to founder on thermodynamic grounds. It most likely

violates the principle of microscopic reversibility, although this can be avoided if the oxidation of NADH is excessively exergonic relative to the synthesis of ATP. However, this would be enormously wasteful in terms of the energy expended to the surroundings, hence is unlikely to have been chosen by evolution. In general, the thermodynamic coupling of an exergonic process to an endergonic one is apparently highly inefficient, as neither of the reactions would go to completion.

Also, the view that ATP is a high energy compound is belied by thermodynamic data, which indicate that ATP hydrolysis is driven by release of phosphate. Many of the reactions involving ATP appear to be *per se* exergonic, so ATP may only be providing part of the activation energy for the reactions (*e.g.* oligopeptide and oligonucleotide synthesis). Under certain conditions, in fact, available thermodynamic data apparently indicate that ATP synthesis (from ADP^{3-} and HPO_4^{2-}) may be marginally exergonic. This may well occur in isolated (possibly hydrophobic) sites, in which the oxidation of NADH to NAD^+ and OH^- leads to high pH levels, which enable the formation of the above highly ionized reactants. This, along with the kinetic stability of ATP, is a likely mode of ATP biosynthesis, and represents a feasible non-equilibrium alternative to the chemiosmotic mode.

APPENDIX

Gibbs free energy, work and entropy

Classical ideas about heat energy, developed at the turn of the nineteenth century, gradually evolved towards a more general concept of energy and its effects on matter. In particular, the development of statistical theories led to the idea that randomness was a fundamental property that was associated with the earlier formulated concept of entropy. In the final decades of the century these ideas crystallized into the well-know concept of the Gibbs free energy (G), defined in terms of the enthalpy (H) and the entropy (S), as in Eqs. (A1) and (A2) (the latter referring to the change in these terms at constant temperature (T).

$$G = H - TS \qquad \text{(A1)}$$

$$\Delta G = \Delta H - T\Delta S \qquad \text{(A2)}$$

$$\mathrm{d}S = \mathrm{d}q_{\mathrm{rev}}/T \qquad \text{(A3)}$$

The enthalpy (H) corresponds closely to the intuitive idea of heat energy. The systemic entropy (S) is formally defined as in Eq. (A3), in terms of the heat absorbed reversibly ($\mathrm{d}q_{\mathrm{rev}}$) by the system at a constant temperature. By Eq. (A3), $T\Delta S$ corresponds to the heat change associated with a change in entropy.

These ideas were originally proposed to explain the working of heat engines, in terms of ideal-gas based systems as theoretical models. In these, apparently, entropy changes corresponded to pressure-volume work under isothermal conditions. However, the advent of the molecular statistical theories of matter and energy required an elaboration of these ideas, leading to the current view of entropy as a measure of randomness.

ΔG is believed to be a measure of both the spontaneity of a process and the work available from it. The criterion for a spontaneous process is $\Delta G < 0$, which corresponds to an overall increase in the entropy of the system and the surroundings [*vide infra* and Eqs. (A2) and (A3)]. ΔG also indicates that the available work from the system does not correspond to the heat change (ΔH) alone, but is also influenced by the systemic entropy change (ΔS).

The enthalpy change corresponds to the systemic heat given up to the surroundings (if $-\Delta H$). This leads to an increase in the entropy of the surroundings, the heat being also available for work therein. However, the interpretation of the change in the systemic entropy is interestingly less straightforward. Currently, it is believed that an increase in the systemic entropy draws in heat from the surroundings, presumably based on the requirements of Eq. (A3). In such a case, the system does not lose any energy to the surroundings, so there is none wasted!

However, the validity of the above interpretation of the systemic entropy change – propounded in many text-books – is unclear. The essential problem with the idea that an increase in the systemic entropy derives from heat drawn from the surroundings, is that the corresponding work performed cannot then be attributed to the system!

An alternative interpretation is based on a reassessment of the concept of entropy itself. It would appear that entropy refers to a type of potential energy that is not part of the internal energy of the system, hence is distinct from its enthalpy. Thus, entropy relates not to the energy content *per se* but to the manner in which the energy is distributed and ordered.

An increase in entropy is thus associated with a release of this distinctive potential energy that increases the disorder within the system. Intriguingly, this is apparently accompanied by the conversion of potential energy to kinetic energy. If there is a substantial volume change (*e.g.* involving a gas), this may manifest as work performed by the system on the surroundings.

It is noteworthy that, in the case of a $-\Delta H$ process, the heat released by the system becomes a part of the surroundings. Its conversion to work would imply that it is performed by the surroundings without involving the system. Thus, the energy involving the enthalpy and entropy changes are distinct and different, and affect the surroundings differently. Importantly, a reaction with a large $+\Delta S$ component would only be expending energy that is anyway unavailable to the surroundings!

The status of the above arguments is currently unclear, but apparently indicates that the origins of entropy are not always straightforward. All the same, it would appear that there do exist fundamental differences between reactions that are driven by a decrease in enthalpy and those driven by an increase in entropy. It also appears that Nature has exploited certain inherent advantages of entropy-driven processes in the case of the Krebs cycle, as discussed below.

In particular, a positive systemic entropy change can be a major contributor to the overall free energy change [Eq. (A2)]. This can offset a modest or even positive enthalpy change ($+\Delta H$), which could lead to the formation of reactive high energy species in the reaction. These would thus be formed in high yields – despite their reactivity – and may be utilized for driving a succeeding endergonic reaction.

Entropy-driven reactions and biological energy coupling

The above arguments indicate a feasible way of coupling a highly exergonic reaction to an endergonic one, without a considerable sacrifice of energy. This can be seen in Fig. (A1), derived from Fig. 2, in which B* and $B^{\#}$ represent high-entropy and high-enthalpy by-products respectively, that are formed exergonically from A. Whilst B* is non-reactive, $B^{\#}$ can take part in an endergonic reaction to form C. A comparison with Fig. 2 shows that the reaction of $B^{\#}$ is less endergonic than that of B.

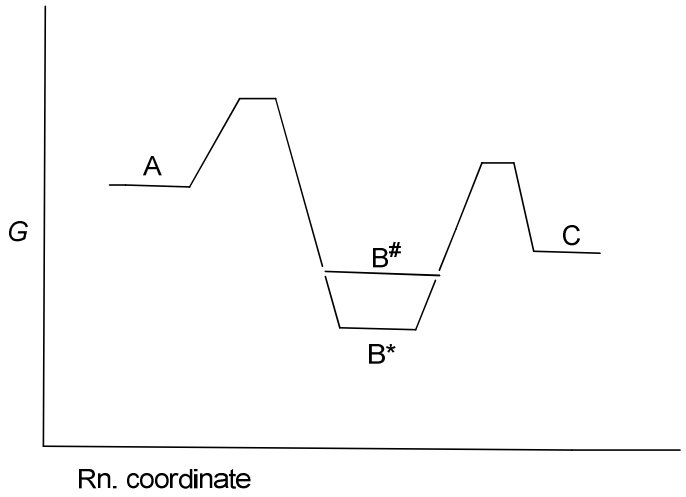


Fig. (A1). *The energy profile for the coupling of an exergonic reaction ($A \rightarrow B^* + B^{\#}$) to an endergonic one ($B^{\#} \rightarrow C$), B* being a non-reactive high-entropic by-product (cf. Fig. 2)*

In light of these arguments, the previously mentioned observation that two of the reactions in the Krebs cycle are oxidative decarboxylations, acquires renewed significance. Thus, the formation of the CO_2 by-product provides the entropic driving force to form species with a relatively low reduction potential, which can transfer an electron irreversibly to NAD^+, leading finally to the formation of NADH. This transfer of free energy to NADH would be enabled by the formation of the non-reactive and high-entropic CO_2 by-product, as argued above.

The above reactions involve the conversion of isocitrate (C_6), *via* α-ketoglutarate (C_5), to succinate (C_4). Each of these steps leads to the formation of one molecule of NADH and one of CO_2. Interestingly, the later conversion of malate (C_4) to oxaloacetate (C_4) also leads to the formation of a molecule of NADH, but no CO_2 as it is a non-degradative oxidation. However, it is possible that this reaction is driven by the subsequent reaction of oxaloacetate with the high energy acetyl-CoA, which enters the Krebs cycle at this stage.

These arguments depend on a particular interpretation of entropy. In fact, they also accord very well indeed with the currently accepted view that an increase in the systemic entropy draws in heat from the surroundings, as this represents no loss to the system (*vide supra*). However, the classical view of entropy is not beyond question, and does not necessarily merge seamlessly with the statistical interpretation. In fact, the classical basis – the isothermal expansion of a heated gas – seems intuitively unviable and may be dubious.[1]

Classically, entropy is defined in terms of the absorption of energy by the system [Eq. (A3)], rather than as a release of potential energy. A much touted application involves melting and evaporation, as in these energy is absorbed isothermally. Thus, the corresponding latent heats are related to the increase in the systemic entropy, although the work done in the phase change is conveniently neglected!

---

[1]Chandrasekhar, S. http://precedings.nature.com/documents/1852/version/1