John Andraos*

Relationships between step and cumulative PMI and E-factors: implications on estimating material efficiency with respect to charting synthesis optimization strategies

Abstract: This report describes mathematical relationships between step and cumulative process mass intensities (PMIs) for synthesis plans, and analogous parameters applied to E-factors. It is shown that both step E-factors and step PMIs are not additive for synthesis plans. It is also shown that a recursive calculation of cumulative PMIs from step PMIs is a rapid method of determining overall PMIs for synthesis plans, though cumulative PMIs are not sufficiently informative as step PMIs or step E-factors to identify bottlenecks in synthesis plans. Illustrations on the use of these metrics to track the material efficiency of published synthesis plans for the pharmaceutical, apixaban, are given as a template example. Advantages and disadvantages of each metric are discussed. A general algorithm to select the most promising candidate synthesis plans considered at the design stage for a given molecular target that most likely satisfy "green" material efficiency criteria is also presented.

Keywords: green chemistry education; green chemistry engineering; sustainability metrics.

Abbreviations

AE	atom economy
BP	by-products
cE	cumulative E-factor
cPMI	cumulative process mass intensity
E	E-factor
m	mass
MW	molecular weight
N	number of reaction steps
PMI	process mass intensity
RS	reaction solvents

*Corresponding author: John Andraos, CareerChem, Toronto, ON M3B 2W4, Canada, e-mail: c1000@careerchem.com

Greek letters

ε	reaction yield
ϕ	mass of input materials

1 Introduction

The use of green chemistry metrics is now a well-established tool used by synthetic organic and process chemists to estimate the material efficiency, environmental impact, and safety-hazard impact of their synthesis plans and chemical processes to a desired target molecule [1, 2]. More importantly, it is a powerful tool in making decisions about which candidate synthesis strategies to choose and ultimately to actualize in the pursuit of "green" syntheses at the early design and planning stages of a synthesis campaign. Even more importantly, it is imperative that whatever algorithms are used to carry out the estimation of those metrics are robust, so that good decisions are made with respect to achieving true synthesis optimization. In this way, chemists will be confident in implementing reliable, reproducible, and intelligible metrics as a routine tool at their disposal in their practice of green chemistry principles in their everyday work.

In this report we wish to present general relationships between step and cumulative process mass intensity (PMI) and general relationships between step and cumulative E-factors (cE) for linear and convergent synthesis plans. Both sets of metrics have been advanced by the pharmaceutical industry as highly relevant to the determination of material efficiency both for "after-the-fact" published plans in the literature and for exploring new synthesis strategies using classical and novel methods at the planning stages *before* any experiments are carried out. Specifically, we develop facile methods of calculating the overall PMI, cumulative PMI, overall E-factor, and cE for synthesis plans. The structure of this report is as follows. We begin by showing the derivation of each method and supply mathematical evidence supporting the connections between the relevant metrics. Next, we illustrate the computation of those metrics for published synthesis plans to the blood anti-coagulant pharmaceutical apixaban **1** [3–8] since this was chosen as a showcase

example by the pharmaceutical industry for achieving a "green" synthesis [9]. Finally, we show how these metrics may be used for ranking the published plans with respect to material efficiency.

2 E-factor and PMI

For any individual chemical reaction, the mass of waste generated may be generally considered to originate from seven sources: by-products (BP) arising as a mechanistic consequence in producing the target product, side products arising as a consequence of competing reactions not leading to the intended target product, unreacted excess reagents, catalysts and ligands, reaction solvents (RS), work-up materials, and purification materials. If m_P is the mass of the desired target product of the reaction, then the total mass of waste and the E-factor are defined according to Eqs. (1) and (2) shown below:

$$m_{waste} = m_{BP} + m_{SP} + m_{XS} + m_{CAT} + m_{RS} + m_{WU} + m_{PU} \tag{1}$$

$$\begin{aligned} E = \frac{m_{waste}}{m_P} &= \frac{m_{BP} + m_{SP} + m_{XS} + m_{CAT} + m_{RS} + m_{WU} + m_{PU}}{m_P} \\ &= \frac{m_{BP}}{m_P} + \frac{m_{SP}}{m_P} + \frac{m_{XS}}{m_P} + \frac{m_{CAT}}{m_P} + \frac{m_{RS}}{m_P} + \frac{m_{WU}}{m_P} + \frac{m_{PU}}{m_P} \\ &= E_{BP} + E_{SP} + E_{XS} + E_{CAT} + E_{RS} + E_{WU} + E_{PU} \end{aligned} \tag{2}$$

We observe from Eq. (2) that the E-factor for the reaction can be decomposed into its contributing E-factors based on the individual sources of the waste. Hence, the overall E-factor for a single chemical reaction is the sum of the contributing E-factors corresponding to each waste source. Now, if we follow a parallel manipulation using the corresponding PMI parameters, then we need to use the general connecting relationship $PMI = E + 1$ [10]. For an individual reaction where m_{input} represents the mass of all input materials used, the expression for PMI can be written as shown in Eq. (3):

$$PMI = \frac{m_{input}}{m_P} = \frac{m_{waste} + m_P}{m_P} = \frac{m_{waste}}{m_P} + 1 = E + 1 \tag{3}$$

Hence, the last line in Eq. (2) can be transformed as shown in Eq. (4):

$$\begin{aligned} PMI - 1 = &(PMI_{BP} - 1) + (PMI_{SP} - 1) + (PMI_{XS} - 1) + (PMI_{CAT} - 1) \\ &+ (PMI_{RS} - 1) + (PMI_{WU} - 1) + (PMI_{PU} - 1) \\ PMI = &PMI_{BP} + PMI_{SP} + PMI_{XS} + PMI_{CAT} + PMI_{RS} + PMI_{WU} \\ &+ PMI_{PU} - 6 \end{aligned} \tag{4}$$

Eq. (4) indicates that the overall PMI for a chemical reaction is *not* the sum of the contributing PMIs as was the case for the E-factors. Rather, it is the sum of the contributing PMIs minus one less than the number of the waste contributing factors. Based on these findings for a single reaction, one might think that the same kind of parallel relationship also holds for sequences of reactions in a synthesis plan; namely, that the overall E-factor is the sum of the step E-factors and that the overall PMI is not the sum of the step PMIs. In fact, for a linear synthesis plan, *neither* the overall E-factor nor the overall PMI is the sum of the corresponding step parameters. We demonstrate what the connecting relationships are between the overall and step parameters for each metric with the following simple logic. For a linear sequence of N steps involving the production of intermediate products P_1, P_2, …, and P_N in steps 1, 2, …, N where all of the preceding intermediate in any given step is committed as a reagent in the next step, we can write the step PMI for step j as shown in Eq. (5):

$$(PMI)_j = \frac{\phi_j}{m_{P_j}}, \tag{5}$$

where ϕ_j is the mass of the input materials used in step j and m_{P_j} is the mass of the intermediate product collected in step j. For the entire synthesis plan covering all N steps, the overall PMI is given by Eq. (6) where we have substituted the step PMI expression given in Eq. (5):

$$\begin{aligned} (PMI)_{1 \to N} &= \frac{\phi_1 + \phi_2 + \cdots + \phi_N - (m_{P_1} + m_{P_2} + \cdots + m_{P_{N-1}})}{m_{P_N}} \\ &= \frac{1}{m_{P_N}} \left[\sum_{j=1}^{N} \phi_j - \sum_{j=1}^{N-1} m_{P_j} \right] \\ &= \frac{1}{m_{P_N}} \left[\sum_{j=1}^{N} m_{P_j} (PMI)_j - \sum_{j=1}^{N-1} m_{P_j} \right] \\ &= \frac{1}{m_{P_N}} \left[m_{P_N} (PMI)_N + \sum_{j=1}^{N-1} m_{P_j} ((PMI)_j - 1) \right] \end{aligned} \tag{6}$$

From Eq. (6), we observe immediately that the overall PMI is generally *not* the sum of the step PMIs. However, the expression *can* collapse to that sum under the very special condition when all the masses of the intermediate products are each identical to the final product of the linear synthesis, i.e. $m_{P_N} = m_{P_j}$ for $j = 1, 2, 3, …, N$. Such a scenario implies a linear sequence comprising consecutive rearrangement reactions where the molecular weight of each intermediate product remains the same and each reaction yield is 100%, with no loss of mass arising from BP or unreacted starting materials along the way. This very rare possibility involving consecutive rearrangement reactions has been

documented before [11]. We can transform the expression given in Eq. (6) to find an analogous expression given in Eq. (7) that relates the overall E-factor for a linear sequence to the step E-factors using the substitution PMI $= E + 1$:

$$E_{1 \to N} + 1 = \frac{1}{m_{P_N}} \left[m_{P_N} (E_N + 1) + \sum_{j=1}^{N-1} m_{P_j} ((E_j + 1) - 1) \right]$$

$$= \frac{1}{m_{P_N}} \left[m_{P_N} (E_N + 1) + \sum_{j=1}^{N-1} m_{P_j} E_j \right]$$

$$E_{1 \to N} = E_N + \frac{1}{m_{P_N}} \sum_{j=1}^{N-1} m_{P_j} E_j \tag{7}$$

Again, we observe that the overall E-factor is generally *not* the sum of the step E-factors, except for the very special case mentioned above. This mathematical evidence therefore shows that the step E-factors for a linear plan are not additive. Close examination of Eq. (7) shows that each of the E-factor parameters is defined with respect to different product masses. Now, if we define the overall mass of waste produced in a linear synthesis plan as the difference between the sum of masses of all input materials used in all reaction steps and the sum of masses of all products produced along the way as shown in Eq. (8):

$$m_{\text{overall waste}} = \sum_{j=1}^{N} \phi_j - \sum_{j=1}^{N} m_{P_j} \tag{8}$$

and then if we divide that mass of overall waste by the mass of the final target product, we obtain, of course, the overall E-factor for the synthesis plan. Furthermore, if we define a *step contributing E-factor*, E_j^*, with respect to the mass of the *final* target product representing the contribution of waste produced from each reaction step to the total waste produced, then we can transform the expression given in Eq. (8) to obtain a relationship that links the overall E-factor to the step contributing E-factors as shown in Eq. (9):

$$\frac{m_{\text{overall waste}}}{m_{P_N}} = \frac{\phi_1 + \phi_2 + \cdots + \phi_N - m_{P_1} - m_{P_2} \cdots - m_{P_N}}{m_{P_N}}$$

$$E_{\text{overall}} = \frac{\phi_1 - m_{P_1}}{m_{P_N}} + \frac{\phi_2 - m_{P_2}}{m_{P_N}} + \cdots + \frac{\phi_N - m_{P_N}}{m_{P_N}}$$

$$E_{\text{overall}} = E_1^* + E_2^* + \cdots + E_N^* \tag{9}$$

In this case, we see that, indeed, the overall E-factor for a linear synthesis plan is the sum of the step contributing E-factors. Note that each step E-factor appearing in Eq. (8) is defined with respect to the mass of the intermediate product produced at that step, whereas the step contributing E-factors appearing in Eq. (9) are all defined with respect to the mass of the final product in the sequence.

3 Cumulative vs. step metrics

We previously reported expressions for cumulative PMI and cE that depended on step PMI and step E-factors, respectively [12, 13]. The recursive relationships for linear sequences are shown in Eqs. (10) and (11):

$$(\text{cPMI})_{1 \to i} = \frac{m_{P_{i-1}}}{m_{P_i}} \left[(\text{cPMI})_{1 \to i-1} - 1 \right] + (\text{PMI})_i \tag{10}$$

$$(cE)_{1 \to i} + 1 = \frac{m_{P_{i-1}}}{m_{P_i}} \left[(cE)_{1 \to i-1} + 1 - 1 \right] + (E)_i + 1$$

$$(cE)_{1 \to i} = \frac{m_{P_{i-1}}}{m_{P_i}} (cE)_{1 \to i-1} + (E)_i \tag{11}$$

where $\dfrac{m_{P_{i-1}}}{m_{P_i}} = \dfrac{\nu_{P_{i-1}} \text{MW}_{P_{i-1}}}{\nu_{P_i} \text{MW}_{P_i} \varepsilon_i}$, MW refers to the molecular weights of the intermediates multiplied by stoichiometric *coefficients*, ν, as appropriate from balanced chemical equations, ε_i is the reaction yield of step i, the $1 \to i$ notation means that the cumulative PMI quantity extends from step 1 to step i, and the counting index goes from $i = 2, 3, \ldots,$ etc. Note that the connecting relation PMI $= E + 1$ was used in obtaining Eq. (11) from Eq. (10). The derivation of Eq. (10) arises as follows. For a three-step plan, from Eq. (5), the step PMIs for steps 1, 2, and 3 are given by Eqs. (12–14).

$$(\text{PMI})_1 = \frac{\phi_1}{m_{P_1}} \tag{12}$$

$$(\text{PMI})_2 = \frac{\phi_2}{m_{P_2}} \tag{13}$$

$$(\text{PMI})_3 = \frac{\phi_3}{m_{P_3}} \tag{14}$$

The cumulative PMI for steps 1 and 2 is given by Eq. (15):

$$(\text{cPMI})_{1 \to 2} = \frac{\phi_1 + \phi_2 - m_{P_1}}{m_{P_2}} = \frac{\phi_1}{m_{P_2}} + \frac{\phi_2}{m_{P_2}} - \frac{m_{P_1}}{m_{P_2}}$$

$$= \frac{\phi_1}{m_{P_1}} \frac{m_{P_1}}{m_{P_2}} + \frac{\phi_2}{m_{P_2}} - \frac{m_{P_1}}{m_{P_2}} = \frac{m_{P_1}}{m_{P_2}} \left[\frac{\phi_1}{m_{P_1}} - 1 \right] + \frac{\phi_2}{m_{P_2}}$$

$$= \frac{m_{P_1}}{m_{P_2}} [(\text{PMI})_1 - 1] + (\text{PMI})_2$$

$$= \frac{m_{P_1}}{m_{P_2}} [(\text{cPMI})_{1 \to 1} - 1] + (\text{PMI})_2 \tag{15}$$

where it is understood that the step PMI for step 1 is identical to the cumulative PMI for step 1. Similarly, the cumulative PMI for steps 1, 2, and 3 is given by Eq. (16):

$$(cPMI)_{1\rightarrow3} = \frac{\phi_1 + \phi_2 + \phi_3 - m_{P_1} - m_{P_2}}{m_{P_3}} = \frac{m_{P_2}(cPMI)_{1\rightarrow2} + \phi_3 - m_{P_2}}{m_{P_3}}$$

$$= \frac{m_{P_2}}{m_{P_3}}\left[(cPMI)_{1\rightarrow2} - 1\right] + (PMI)_3 \qquad (16)$$

It is readily apparent from the emerging pattern of Eqs. (15) and (16) that Eq. (10) is verified for general step i. For a linear sequence consisting of N steps, the cumulative PMI, $(cPMI)_{1\rightarrow N}$ is identical to the overall PMI that would be obtained from Eq. (6). It is also understood from this analysis that all intermediate products produced along the linear chain are limiting reagents in each reaction, so the molar stoichiometric ratio is always 1. When we are dealing with a convergent synthesis, the same kind of analysis is done on the convergent branch from the first step of that branch to the final step of that branch. A corresponding cumulative PMI will therefore be obtained up to the terminal intermediate of the convergent branch. At the convergent step, there will be a reaction between an intermediate product from the main branch and the terminal product of the convergent branch along with any other reagents. In a properly designed synthesis, the intermediate product from the main branch will be the limiting reagent, but the terminal intermediate product from the convergent branch may be used in excess. Hence, the expression for the cumulative PMI from step 1 up to the convergent step along the main branch will have an extra term included that accounts for the contribution arising from the convergent branch. For example, if we have a convergent synthesis plan having a convergent branch consisting of three steps leading up to intermediate P_{3*} and the convergent step occurs at step 5 along the main branch, then the cumulative PMI from step 1 to step 5 is given by Eq. (17):

$$(cPMI)_{1\rightarrow5} = \frac{m_{P_4}}{m_{P_5}}[(cPMI)_{1\rightarrow4} - 1] + (PMI)_5$$

$$+ \frac{m_{P_{3*}}}{m_{P_5}}[(cPMI)_{1*\rightarrow3*} - 1] \qquad (17)$$

The first two terms in Eq. (17) arise from the normal recursion formula applied to the main branch and the last term pertains to the contribution arising from the convergent branch. The expressions for the two mass ratios of intermediate products are shown in Eqs. (18) and (19):

$$\frac{m_{P_4}}{m_{P_5}} = \frac{\nu_{P_4} MW_{P_4}}{\nu_{P_5} MW_{P_5} \varepsilon_5} \qquad (18)$$

$$\frac{m_{P_{3*}}}{m_{P_5}} = \frac{\nu_{P_{3*}} MW_{P_{3*}}}{\nu_{P_5} MW_{P_5} \varepsilon_5} \times \frac{\text{actual moles}_{P_{3*}}}{\text{stoichiometric moles}_{P_{3*}}} \qquad (19)$$

where the mole ratio in Eq. (19) is a number that has a value greater than 1. Appendix 1 contains a more complex example of a convergent plan drawn in the form of a synthesis tree diagram [14, 15] containing five branches including the main branch (see Figure S1). From that diagram, it is possible to write out all the cumulative PMI expressions sequentially from the first step all the way to the final step by inspection [see Eqs. (S1–S17)].

4 Application to syntheses of apixaban

We evaluated the material efficiency of various literature-documented syntheses of the blood anti-coagulant apixaban **1** as a demonstration of the mathematical relationships presented in this work. Six synthesis plans were considered including two from Bristol-Myers Squibb [3, 4], one from an academic lab in China [5], and three from generic drug companies [6–8]. All of them follow the same convergent synthesis strategy as evidenced by the common target bond mapping shown in Figure 1, which highlights the construction bonds made in the product chemical structure. The mapping for the pyrazolo[3,4-c]pyridin-7-one ring frame can be encoded as $[(2+1+1+1)+(5+1)]$.

Table 1 summarizes the essential materials efficiency metrics for all six plans which are listed in ascending order of total PMI. These metrics were calculated using our previously reported REACTION and SYNTHESIS
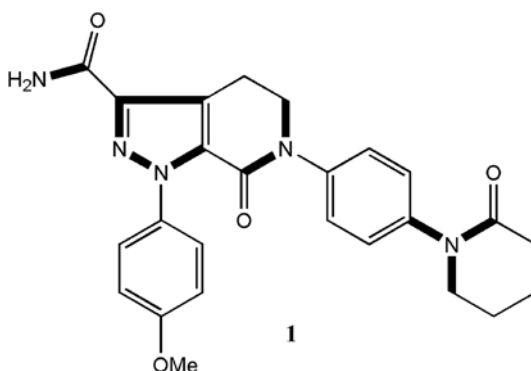

Figure 1: Target bond mapping for all published synthesis plans to apixaban, **1**.

Table 1: Summary of green metrics for various synthesis plans to apixaban.

Plan	# Steps	Overall Y (%)[a]	Overall AE (%)	E-kernel	E-excess	E-solvent	E-cat	E-workup	E-purif	E-total	PMI-total
Bristol-Myers 2006	7	37.4	20.8	10.32	7.72	80.65	1.12	116.85	27.2	243.86	244.86
Optimus 2017	9	16.7	18.4	18.63	15.90	117.40	2.46	214.10	9.41	377.90	378.90
Jiang-Ji 2013	7	26.8	18.4	11.15	16.84	104.15	0.05	220.37	29.53	382.09	383.09
MSN 2015	8	16.3	19.2	24.29	28.9	126.45	0.63	308.40	8.78	497.45	498.45
Bristol-Myers 2003	7	25.6	18.5	16.06	19.15	128.22	0.80	460.73	0	624.96	625.96
Dr. Reddy 2017	7	22.2	19.8	19.86	24.24	239.37	0.86	442.27	129.36	855.96	856.96

[a]Calculated along the longest branch. Green color indicates the best performance; red color indicates the worst performance.
PMI, Process mass intensity.

spreadsheets [11]. The best metric parameters are highlighted in green and the worst ones are shown in red. This ranking shows that the Bristol-Myers Squibb 2006 plan has the best overall mass efficiency profile, having the best performances with respect to the highest overall yield and overall atom economy, and the lowest values for E-kernel, E-excess, E-solvent, E-workup, E-total, and PMI-total. That plan corresponds to the one reported in a recent paper on cumulative PMI calculations [9]. The Jiang-Ji plan had the lowest E-cat contribution. The separation of E-total contributions assists in pinpointing each plan's strengths and weaknesses. The main liabilities impacting the last place Dr. Reddy plan are a high RS consumption and a high work-up auxiliary material consumption. The significant improvement in the Bristol-Myers Squibb 2006 plan over their earlier 2003 plan is mainly due to an 11% gain in overall yield and reductions in E-solvent and E-workup by 59% and 294%, respectively. It is not surprising that most of the potential gains in "greening up" synthesis procedures arise from solvent and other auxiliary material consumption, since these account for at least 75% of all the input mass used in the chemical reactions. As a consequence, pharmaceutical companies have focused their efforts to reduce auxiliary material consumption as a primary strategy to achieve green syntheses of their products. We can drill deeper to find the precise reaction steps in each plan that are responsible for any mass inefficiencies by examining the shapes of the corresponding radial pentagon diagrams and correlating them to which of the following four parameters is responsible: reaction yield, step AE, step SF (stoichiometric factor representing excess energy consumption), and MRP (material recovery parameter representing auxiliary material consumption). Figure 2 shows the radial pentagon diagrams for the Bristol-Myers Squibb 2006 apixaban plan. By visual inspection, that plan has the following liabilities: (a) low AE for steps 1 (amidation-α,α′-dichlorination sequence) and 2b (diazotization-imination sequence) and

(b) high SF for steps 2b and 6 (final carboxyamination). Applying the same analysis to the other plans we have the following results. For the Optimus 2017 plan we have the following liabilities: (a) low AE for steps 2 (cyclic amination), 3 (α,α′-dichlorination), 5 (Heck coupling), and 5b (diazotization-imination sequence), and (b) high SF for steps 5, 5b, and 8 (final carboxyamination). For the Jiang-Ji 2013 plan we have the following liabilities: (a) low AE for steps 1 (cyclic amidation-amination sequence), 2 (α,α′-dichlorination-morpholine substitution sequence), 3 (aromatic nitro group reduction with sodium sulfide), 4 (cyclic amidation-amination sequence), and 4b (diazotization-imination sequence), and (b) high SF for steps 1, 2, and 4b. For the MSN 2015 plan we have the following liabilities: (a) low reaction yields for steps 2 (cyclic amination-α,α′-dichlorination-chlorine elimination sequence) and 4b (diazotization-imination sequence), (b) low AE for steps 1 (acid chloride formation-amidation sequence), 2, 4 (Heck coupling), and 4b, and (c) high SF for steps 4b, 5 (pyrazole formation), and 6 (carboxyamination). For the Bristol-Myers Squibb 2003 plan we have the following liabilities: (a) low reaction yields for steps 2 (α,α′-dichlorination-morpholine substitution sequence) and 5 (final carboxyamination), (b) low AE for steps 1 (amidation-cyclic amination sequence), 2, 3 (Heck coupling), and 2b (imination), and (c) high SF for steps 2, 4 (pyrazole formation), and 5. Finally, for the Dr. Reddy 2017 plan we have the following liabilities: (a) low reaction yield for step 3 (pyrazole formation), (b) low AE for steps 1 (α,α′-dichlorination), 2 (chlorine elimination), 2b (diazotization-imination sequence), 3, 3b (amidation-cyclic amination sequence), and 4 (Heck coupling), and (c) high SF for steps 2b, 3, and 4.

Based on these findings, we can also identify the most optimum step PMI values corresponding to the target bonds made according to the target bond map given in Figure 1. These are summarized in Figure 3 where it appears that the second-ranked Optimus 2017 plan has
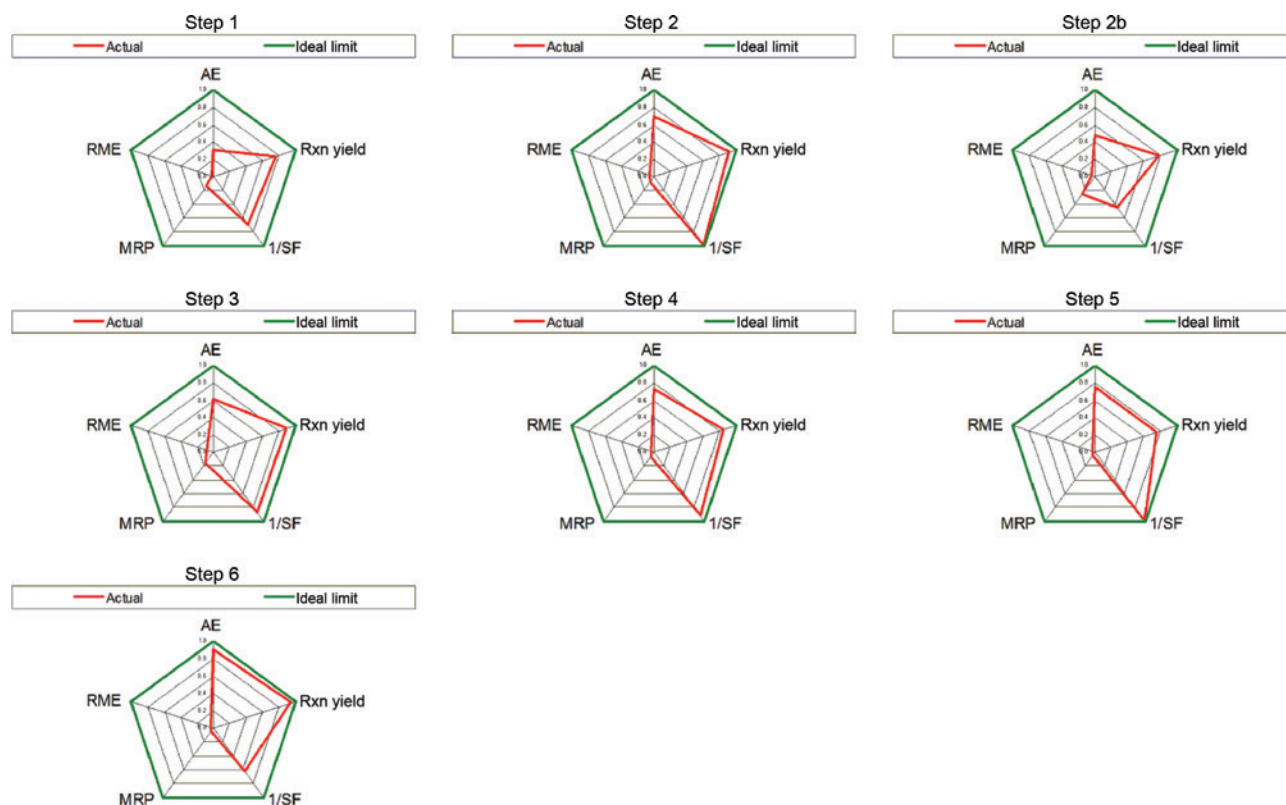
Figure 2: Radial pentagons for Bristol-Myers Squibb 2006 plan for apixaban.

the lowest step PMI values for four out of the seven kinds of target bond made, though from its E-factor contributions shown in Table 1, it has disadvantages in all but the E-purification category compared to the best forming BMS 2006 plan with the lowest overall PMI. From a synthesis strategy point of view, a better discriminator of material efficiency than step PMI by itself is the mass of sacrificial reagents used in a plan per kg of target product. Examples of sacrificial reagents are acids or bases, reagents that introduce or remove protecting groups, reagents that introduce or exchange functional groups to improve leaving group ability, redox reagents that correct oxidation states of key atoms in order to make them more electrophilic or nucleophilic, as the case may be for bond formation in later reaction steps, etc. where none of their atoms are incorporated at all in the structure of the final target product. Such a parameter offsets any effects of auxiliary material consumption, which are embedded in step PMI values, and focuses more on the waste produced as a direct consequence of making target bonds. Thus, in order to make 1 kg of apixaban, the corresponding masses of sacrificial reagents used according to the plans examined given in ascending order are as follows: (a) BMS 2006, 12.43 kg; (b) Optimus 2017, 24.01 kg; (c) Jiang-Ji 2013, 24.33 kg; (d) BMS 2003, 27.67 kg; (e) Dr. Reddy 2017, 35.56 kg; and (f) MSN 2015, 36.26 kg. This ranking parallels that

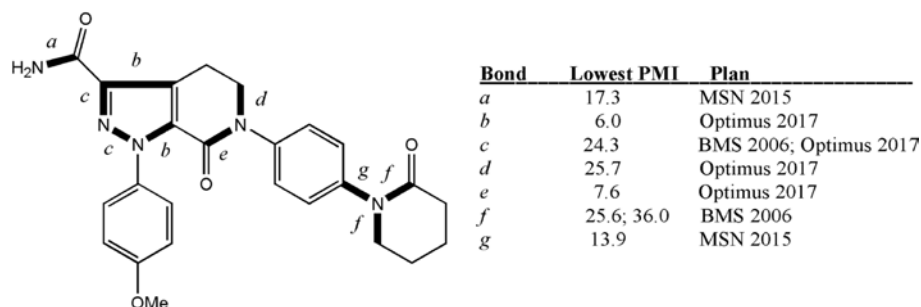| Bond	Lowest PMI	Plan
a	17.3	MSN 2015
b	6.0	Optimus 2017
c	24.3	BMS 2006; Optimus 2017
d	25.7	Optimus 2017
e	7.6	Optimus 2017
f	25.6; 36.0	BMS 2006
g	13.9	MSN 2015

Figure 3: Target bond mapping for apixaban showing optimal step PMI performances according to specific target bonds made.

found for the overall PMI shown in Table 1 for the top three ranking plans; however, the sacrificial reagent ranking for the next three plans are offset by their significant and overwhelming auxiliary material consumptions.

For comparison, we also summed the step E-factors for each plan and obtained the following values: 197.81 (Bristol-Myers, 2006); 208.72 (Optimus); 338.30 (Jiang-Ji); 274.18 (MSN); 419.51 (Bristol-Myers, 2003); and 688.55 (Dr. Reddy). We can see that in all cases, the values obtained by summing the step E-factors are noticeably less than the true values of E-total shown in the 11th column of Table 1. The additive values are found to be between 12% and 45% lower than the true overall E-factor values. This shows that in this specific case, the summing of step E-factors results in a significant underestimation of the overall E-factor. However, this result does not always hold true all the time, as we demonstrate that there are other cases when the sum of the step E-factors may exceed the true overall E-factor. In Appendix 2 we show as examples all possible case scenarios for general two- and three-step synthesis. For longer syntheses it is difficult to make predictions as to the relative outcome of the sum of E-factors vs. the true overall E-factor, because the number of possible scenario cases increases dramatically making the task impractical. *What is certain is that the sum of step E-factors does not equal the true overall E-factor for a given synthesis plan.*

Following our developed cumulative PMI calculator spreadsheet based on the recursive relation given in Eq. (10) (see Supporting Information) we were able to easily obtain the cumulative PMI values at each reaction step for all synthesis plans examined. The input data include the reaction yields, the molecular weights of intermediate products and final product, the mole ratios of terminal intermediate products from the convergent branches, and the step PMIs. The step PMIs were obtained from the REACTION spreadsheets applied to each reaction step using the reported experimental masses of materials used. The results are reported in Tables 2–4 corresponding to complete, kernel, and kernel plus excess cumulative PMI values, respectively. The three levels of calculation correspond to how the step PMI values are determined. The complete calculations correspond to all materials used, the kernel calculations correspond to only the atom economy and reaction yield, and the kernel plus excess reagent contribution corresponds to the atom economy, reaction yield, and stoichiometric factor as shown in Eqs. (20)–(22):

Table 2: Summary of cumulative complete process mass intensity (PMI) values for various synthesis plans to apixaban.

Step j	BM-2006	Optimus-2017	Jiang-Ji-2013	MSN-2015	BM-2003	Dr. Reddy-2017
1	45.92	7.62	36.93	10.35	70.23	65.82
2	75.31	35.40	101.29	107.46	167.16	289.54
3	78.81	78.61	151.55	124.80	288.44	399.80
4	95.52	106.71	214.04	190.78	354.82	579.92
5	183.54	189.87	308.68	224.37	625.64	857.06
6	245.01	161.64	382.83	336.47		
7		269.01		498.35		
8		379.30				

Yellow color indicates final global PMI values.

Table 3: Summary of cumulative kernel process mass intensity (PMI) values for various synthesis plans to apixaban.

Step j	BM-2006	Optimus-2017	Jiang-Ji-2013	MSN-2015	BM-2003	Dr. Reddy-2017
1	4.30	1.79	2.68	1.71	2.59	4.00
2	5.69	3.45	5.36	6.48	6.06	6.28
3	6.11	5.64	7.59	7.78	10.06	11.43
4	6.20	6.97	8.07	11.99	10.87	13.86
5	9.90	9.88	10.29	13.64	16.56	18.65
6	11.17	9.19	12.05	19.79		
7		14.51		24.73		
8		19.63				

Yellow color indicates final kernel PMI values.

$$\text{step PMI}_{\text{complete}} = \left[(\text{Yield})(\text{AE})\left(\frac{1}{\text{SF}}\right)(\text{MRP})\right]^{-1} \quad (20)$$

$$\text{step PMI}_{\text{kernel}} = [(\text{Yield})(\text{AE})]^{-1} \quad (21)$$

$$\text{step PMI}_{\text{kernel+excess}} = \left[(\text{Yield})(\text{AE})\left(\frac{1}{\text{SF}}\right)\right]^{-1} \quad (22)$$

where SF is the stoichiometric factor [16], and MRP is the material recovery parameter accounting for all auxiliary materials used [16].

The highlighted values in yellow shown in Tables 2–4 correspond to the overall cumulative PMIs for each plan under the three regimes of calculation. In all cases, the results of the final cumulative PMIs agreed with the overall PMIs obtained from the SYNTHESIS spreadsheet analyses. For example, the highlighted values shown in Table 2 agree with those given in the last column of Table 1. Table 5 shows the rankings of all plans based on these overall cumulative PMI values. We can see that the Bristol-Myers Squibb 2006 plan ranks first in *all* three regimes, whereas the ranking order changes for the other plans. This actually is an excellent test for robustness of "green" material efficiency for a synthesis. If a plan consistently ranks first under these three regimes, then the

probability that it represents the "greenest" plan to a given target is very high. Ranking first under the kernel regime means that the plan likely has the highest atom economy and overall yield; ranking first under the kernel plus excess regime means that it also consumes the least excess reagents; and finally, ranking first under the complete regime means that it also consumes the least auxiliary materials in terms of RS, work-up materials, and purification materials.

A close examination of the rankings of the other plans shown in Table 5 shows some significant changes depending on the calculation regime. For example, the Optimus plan is ranked #2 under the complete regime but drops to #5 under the kernel regime. This is consistent with its auxiliary material consumption being close to that of the #1 plan, and its low ranking overall yield – about half that of the #1 plan. Apart from the winning plan, the Jiang-Ji and MSN plans show little variation in ranking among the three regimes. It is clear from this example that true overall synthesis optimization is achieved when optimization in all three regimes is orchestrated in the same direction. We believe and emphasize that this criterion is the best and most significant indicator that controls material efficiency optimization of synthesis plans, and therefore the achievement of comparatively "green" synthesis plans. We also point out that in order to drill down to discover where the

Table 4: Summary of cumulative kernel plus excess process mass intensity (PMI) values for various synthesis plans to apixaban.

Step j	BM-2006	Optimus-2017	Jiang-Ji-2013	MSN-2015	BM-2003	Dr. Reddy-2017
1	6.24	2.14	4.46	2.10	4.19	5.61
2	8.14	5.17	11.03	9.56	14.16	9.18
3	10.44	8.45	15.55	11.25	22.79	27.38
4	10.25	10.38	16.48	17.77	22.41	33.57
5	16.26	15.75	24.66	26.34	36.22	45.12
6	19.04	15.76	28.95	43.35		
7		25.50		54.19		
8		35.56				

Yellow color indicates final kernel plus excess PMI values.

Table 5: Summary of cumulative process mass intensity (PMI) ranks for various synthesis plans to apixaban.

Plan	Complete cPMI	Rank (complete)	Kernel cPMI	Rank (kernel)	Kernel+excess cPMI	Rank (kernel+excess)
BM-2006	245.01	1	11.17	1	19.04	1
Optimus-2017	379.3	2	19.63	5	35.56	3
Jiang-Ji-2013	382.83	3	12.05	2	28.95	2
MSN-2015	498.35	4	24.73	6	54.19	6
BM-2003	625.64	5	16.56	3	36.22	4
Dr. Reddy-2017	857.06	6	18.65	4	45.12	5

material throughput bottlenecks are in any given plan, it is insufficient to look only at the overall E-factor, overall PMI, overall AE, and overall yield values as guides, or for that matter cumulative counterparts of those parameters. Two important guides are the partitioning of the E-factors as shown in Table 1, and radial pentagon diagrams for each reaction step in a synthesis as shown in the REACTION spreadsheets for each plan and discussed elsewhere [12]. We determined the precise liabilities of each synthesis plan for apixaban by following this procedure as described in our earlier discussion. We note that the computation of overall PMI via a recursive cumulative PMI calculation is indeed facile. However, the tracking of the incremental increases in PMI does not shed enough light on spotting the bottlenecks in a synthesis plan in comparison to tracking individual reaction step PMIs in conjunction with their radial pentagon diagrams which show how the partitioned step metrics come together to result in the corresponding step PMI values.

Though it appears that synthesis optimization of apixaban using the existing synthesis strategy has reached a plateau, it does not mean that entirely new strategies cannot be considered that would further drive optimization forward. What has been achieved is comparatively the "greenest" synthesis plan within the constraint of a particular design synthesis strategy, as shown by the common target bond mapping shown in Figure 1. We demonstrate that it is possible to use the kernel green metrics of the reaction yield and atom economy applied to new synthesis schemes employing new synthesis design strategies to the same target molecule. Once a scheme is conjectured, its step and overall atom economies can be determined. Estimates of reaction yield performance
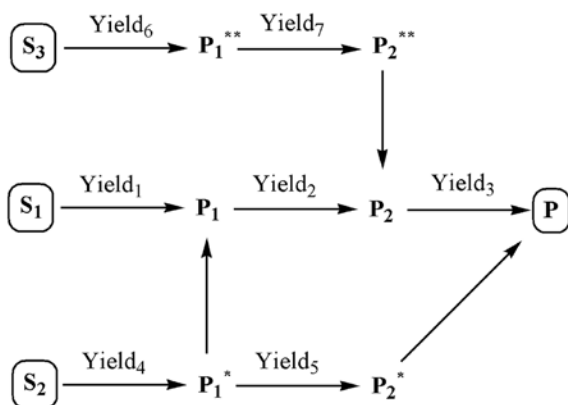
at each step in a conjectured scheme may be made via literature searches where minimum and maximum yield ranges may be obtained. Selecting maximum yield values based on literature evidence coupled with atom economy metrics will yield best-case scenario step PMI and cumulative PMI computations under the kernel regime. These results may be compared with those from the documented literature plans computed under the same regime. In addition, these results may then be judged to see if any of the conjectured schemes have any
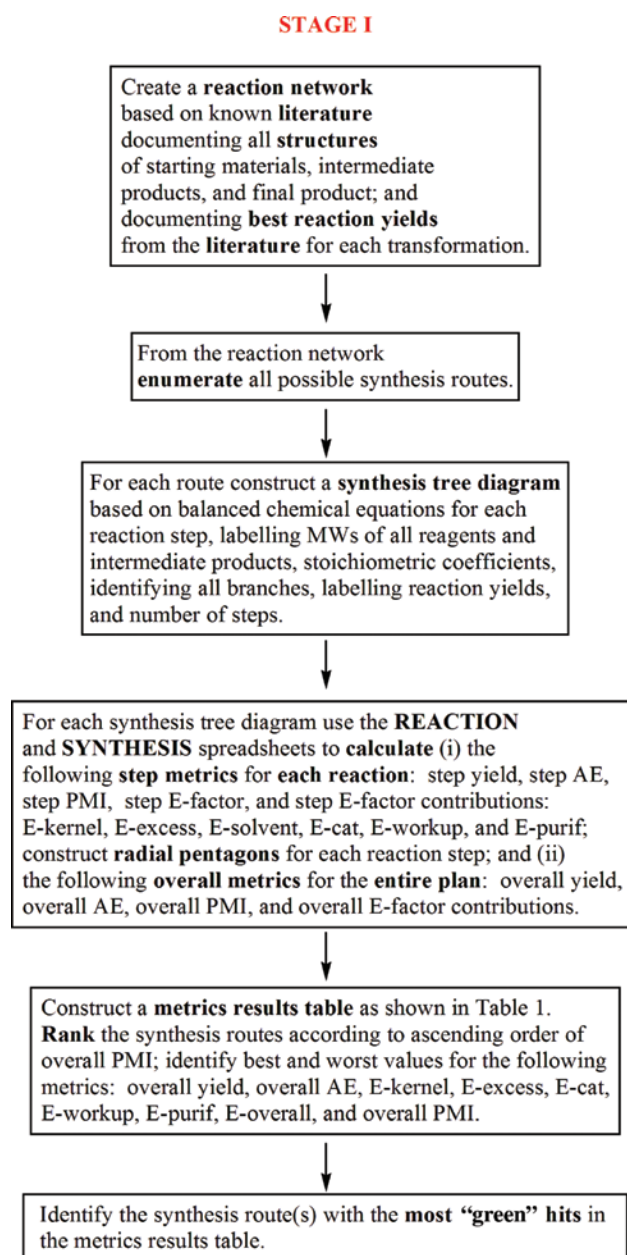

Figure 4: Example reaction network for the synthesis of product P via various synthesis routes from various starting materials showing the best available literature yields for each transformation.

Figure 5: Flowchart for Stage I of the metrics analyses for documented literature synthesis plans.

STAGE II

Conjecture **new synthesis routes** to a target molecule structure with a view to obtain a different target bond mapping from published routes.

↓

For each route draw a **synthesis tree diagram** including all parameters as in Stage I.

↓

Estimate the **lowest and highest reaction yields** for each step in each route using literature evidence.

↓

For each route calculate the following **metrics at the kernel level**: overall AE, min. overall yield, max. overall yield, min. kernel overall PMI, max. kernel overall PMI

↓

Rank the synthesis routes according to ascending order of max. overall PMI; identify best and worst values for the following metrics: overall yield and overall AE.

↓

Compare the results of the ranking with those from STAGE I at the kernel level.

↓

Identify any new conjectured routes that show improvements in the kernel metrics.

↓

For each route **identify the following drawbacks:**
(i) availability of starting materials
(ii) cost of starting materials
(iii) potential for unwanted side reactions
(iv) potential for safety and hazardous materials.

↓

Identify the best candidate conjectured routes satisfying the least drawbacks and the best kernel metrics profile.

Figure 6: Flowchart for Stage II of the metrics analyses for conjectured synthesis plans.

merit for further investigation pending criteria such as the availability and cost of starting materials, environmental impact issues with respect to input materials and BP produced, safety and hazard issues pertaining to the use of any reagent or the handling of waste reaction BP, opportunities for recycling and/or reclaiming, limitations on the transport of reactive reagents, anticipation of unwanted side reactions occurring, anticipated ease of purifying intermediates derived from anticipated side reactions, e.g. formation of isomers, etc.

5 Algorithm for finding the optimum material efficient synthesis plan

Based on our ongoing research spanning 15 years on the study of quantitative parameterization of synthesis and green metrics development, we are able to put forward the following algorithmic protocol for finding the *optimum material efficient synthesis route* to a desired known target molecule. Two main stages are involved: (i) in stage I, a thorough literature search of existing plans is carried out with a view to create a complete reaction network as shown by the generalized example given in Figure 4; and (ii) in stage II, based on the findings of stage I, a brainstorming of ideas is summoned to come up with new synthesis strategies not documented in the literature for the same target molecule. Stage I involves a comprehensive analyses of known synthesis plans from properly documented experimental procedures using step and overall reaction yield, atom economy, E-factor, E-factor partitions, and PMI metrics conducted under the complete and kernel regimes. Stage II involves inventing new synthesis strategies leading to different target bond maps for the same target molecule and assessing their potential for material efficiency "greenness" using the same suite of metrics as Stage I, but under the kernel regime.

In the case of the apixaban plans studied in this work, the corresponding reaction network summarizing all intermediates involved in all routes is shown in Figure 5. Figures 6 and 7 show flowcharts describing the logical flow of steps involved in stages I and II. We followed both sets of protocols in presenting our findings on the synthesis of apixaban in this work.

6 Conclusions

In this work, we have established the mathematical connections between step and cE and PMI metrics with a view to facilitate computation of these quantities for synthesis plans. The statement that the sum of step E-factors for a synthesis plan does not equal the overall E-factor has been confirmed. Similarly, the sum of step PMIs for a synthesis plan does not equal the overall PMI. We also presented a simple computation of cumulative PMI and produced a simple calculator that automates the calculation for both linear and convergent plans. We have demonstrated the utility of a metrics-guided algorithm to both comprehensively assess literature synthesis plans
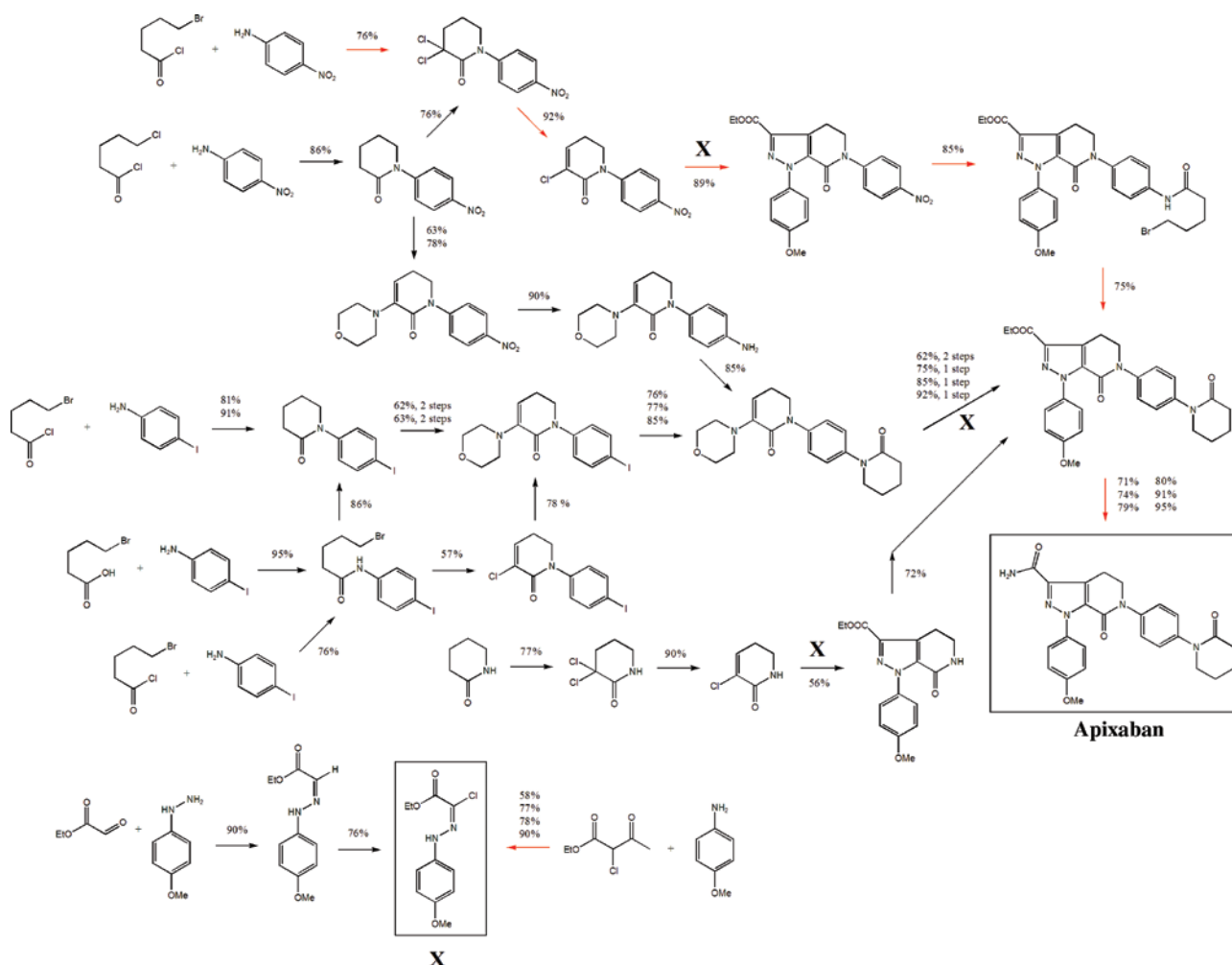
Figure 7: Reaction network for the syntheses of apixaban. Steps shown with red arrows refer to the best performing BMS 2006 route.

to a given target molecule and to suggest new synthesis strategies for the same product using the synthesis of the pharmaceutical apixaban as an example. We have also shown that the total mass of sacrificial reagents per kg of the target product is a more insightful parameter than overall PMI alone in assessing intrinsic synthesis strategy efficiency in the use of target bond forming or construction reactions over concession reactions, since it is not skewed by auxiliary material consumption. We believe that the quantitative parameterization of synthesis performance of individual reactions and synthesis plans (both linear and convergent) by green metrics at the level of material efficiency is now a completely solved mathematical problem and does not require further investigation or invention of new metrics or algorithms. The only outstanding problem is deciding how far back one needs to trace intermediates and reagents to a set of progenitor "starting materials" in order to conduct a materials efficiency metrics analysis of a given synthesis plan. This

will have an effect on the magnitude of the overall PMI calculated for a synthesis plan. The inclusion of PMIs of syntheses of catalysts and other intermediates to the calculation of overall PMIs for the synthesis of a target product will obviously have an augmenting effect, as has been noted before [9, 12, 13, 17]. The key question to address, therefore, is what is the standard progenitor set of reagents and starting materials from which all synthesis plans are evaluated with respect to calculating the overall PMI in a systematic and unbiased way without conferring any undue advantage or disadvantage to any given plan or process. One of the tasks of the recently convened IUPAC Metrics for Green Syntheses (MGS-1) Project is to tackle this question [13]. We hope that this work will put chemists on a confident path to optimize their own synthesis plans with the goal of practicing green chemistry principles and making intelligent decisions about route selection based on material efficiency in a simple, logical, robust, and reliable manner.

Appendix 1: Supplementary mathematical information.
Convergent steps: step 3, step 7, step 9
Cumulative PMI expressions:
(i) Main branch

$$(cPMI)_{1\to1} = (PMI)_1 \tag{S1}$$

$$(cPMI)_{1\to2} = \frac{m_{P_1}}{m_{P_2}}[(cPMI)_{1\to1} - 1] + (PMI)_2 \tag{S2}$$

$$(cPMI)_{1\to3} = \frac{m_{P_2}}{m_{P_3}}[(cPMI)_{1\to2} - 1] + (PMI)_3$$
$$+ \frac{m_{P_{2^*}}}{m_{P_3}}[(cPMI)_{1^*\to2^*} - 1] + \frac{m_{P_{2^{**}}}}{m_{P_3}}[(cPMI)_{2^{**}\to2^{**}} - 1] \tag{S3}$$

$$(cPMI)_{1\to4} = \frac{m_{P_3}}{m_{P_4}}[(cPMI)_{1\to3} - 1] + (PMI)_4 \tag{S4}$$

$$(cPMI)_{1\to5} = \frac{m_{P_4}}{m_{P_5}}[(cPMI)_{1\to4} - 1] + (PMI)_5 \tag{S5}$$

$$(cPMI)_{1\to6} = \frac{m_{P_5}}{m_{P_6}}[(cPMI)_{1\to5} - 1] + (PMI)_6 \tag{S6}$$

$$(cPMI)_{1\to7} = \frac{m_{P_6}}{m_{P_7}}[(cPMI)_{1\to6} - 1] + (PMI)_7$$
$$+ \frac{m_{P_{6^{***}}}}{m_{P_7}}[(cPMI)_{4^{***}\to6^{***}} - 1] \tag{S7}$$

$$(cPMI)_{1\to8} = \frac{m_{P_7}}{m_{P_8}}[(cPMI)_{1\to7} - 1] + (PMI)_8 \tag{S8}$$

$$(cPMI)_{1\to9} = \frac{m_{P_8}}{m_{P_9}}[(cPMI)_{1\to8} - 1] + (PMI)_9$$
$$+ \frac{m_{P_{8^{****}}}}{m_{P_9}}[(cPMI)_{7^{****}\to8^{****}} - 1] = (PMI)_{total} \tag{S9}$$

(ii) Branch 2

$$(cPMI)_{1^*\to1^*} = (PMI)_{1^*} \tag{S10}$$

$$(cPMI)_{1^*\to2^*} = \frac{m_{P_{1^*}}}{m_{P_{2^*}}}[(cPMI)_{1^*\to1^*} - 1] + (PMI)_{2^*} \tag{S11}$$

(iii) Branch 3

$$(cPMI)_{2^{**}\to2^{**}} = (PMI)_{2^{**}} \tag{S12}$$

(iv) Branch 4

$$(cPMI)_{4^{***}\to4^{***}} = (PMI)_{4^{***}} \tag{S13}$$

$$(cPMI)_{4^{***}\to5^{***}} = \frac{m_{P_{4^{***}}}}{m_{P_{5^{***}}}}[(cPMI)_{4^{***}\to4^{***}} - 1] + (PMI)_{5^{***}} \tag{S14}$$

$$(cPMI)_{4^{***}\to6^{***}} = \frac{m_{P_{5^{***}}}}{m_{P_{6^{***}}}}[(cPMI)_{4^{***}\to5^{***}} - 1] + (PMI)_{6^{***}} \tag{S15}$$

(v) Branch 5

$$(cPMI)_{7^{****}\to7^{****}} = (PMI)_{7^{****}} \tag{S16}$$

$$(cPMI)_{7^{****}\to8^{****}} = \frac{m_{P_{7^{****}}}}{m_{P_{8^{****}}}}[(cPMI)_{7^{****}\to7^{****}} - 1] + (PMI)_{8^{****}} \tag{S17}$$

Appendix 2: Case scenarios for a ***two-step*** plan with respect to the sum of the step E-factors and the overall E-factor.
For a two-step plan given by step 1: $A + B \to P_1$, and step 2: $P_1 + C \to P_2$ the overall E-factor in terms of the step E-factors E_1 and E_2 is given by Eq. (S18).

$$E_T = \frac{m_{P_1}}{m_{P_2}}E_1 + E_2 \tag{S18}$$

Case I
When $\dfrac{m_{P_1}}{m_{P_2}} = 1$, then $E_T = E_1 + E_2$.
Case II
When $\dfrac{m_{P_1}}{m_{P_2}} > 1$, then $E_T = E_1 > E_2$.
Case III

When $\dfrac{m_{P_1}}{m_{P_2}} < 1$, then $E_T < E_1 + E_2$.

Case scenarios for a *three-step* plan with respect to the sum of the step E-factors and the overall E-factor.

For a three-step plan given by step 1: $A + B \rightarrow P_1$, step 2: $P_1 + C \rightarrow P_2$, and step 3: $P_2 + D \rightarrow P_3$ the overall E-factor in terms of the step E-factors E_1, E_2, and E_3 is given by Eq. (S19).

$$E_T = \frac{m_{P_1}}{m_{P_3}} E_1 + \frac{m_{P_2}}{m_{P_3}} E_2 + E_3 \qquad (S19)$$

Figure S2 summarizes the nine possible case scenarios for a three-step plan with respect to the magnitude of the mass ratio of intermediates whether they are equal to 1, greater than 1, or less than 1. Seven of the nine cases yield definite conclusions with respect to the relative magnitudes of the true overall E-factor and the sum of the step E-factors.

References

[1] Lapkin A, Constable D, Eds., *Green Chemistry Metrics–Measuring and Monitoring Sustainable Processes*, Wiley: Chichester, 2008.

[2] Constable DJC, Jiménez-González C, Eds., *Handbook of Green Chemistry–Green Metrics*, Wiley-VCH: Weinheim, 2018, Vol. 11.

[3] Zhou J, Oh LM, Ma P, Li HY, Confalone P. US2003181466 (Bristol-Myers Squibb Company, 2003).

[4] Shapiro R, Rossano LT, Mudryk BM, Cuniere N, Obelholzer M, Zhang H, Chen BC. US2006069258 (Bristol-Myers Squibb Company, 2006).

[5] Jiang J, Ji Y. *Synth. Commun*. 2013, 43, 72–79.

[6] Rajan ST, Eswaraiah S, Venkatesh M. US2015353541 (MSN Laboratories Limited, 2015).

[7] Nevuluri NR, Rapolu RK, Iqbal J, Kandagatla B, Sen S, Dahanukar VH, Oruganti S. *Monatsh. Chem*. 2017, 148, 1477–1482.

[8] Reddy D, Reddy S, Rane DR, Velivela SR. WO2017187245 (Optimus Drugs (P) Limited, 2017).

[9] Li J, Albrecht J, Borovika A, Eastgate MD. *ACS Sustainable Chem. Eng*. 2018, 6, 1121–1132.

[10] Jiménez-González C, Ponder CS, Broxterman QB, Manley JB. *Org. Process Res. Dev*. 2011, 15, 912–917.

[11] Andraos J, Hent A. *J. Chem. Educ*. 2015, 92, 1831–1839.

[12] Andraos J. *ACS Sustainable Chem. Eng*. 2016, 4, 1917–1933.

[13] Andraos J. *ACS Sustainable Chem. Eng*. 2018, 6, 3206–3214.

[14] Andraos J. *Org. Process Res. Dev*. 2006, 10, 212–240.

[15] Andraos J, Hent A. *J. Chem. Educ*. 2015, 92, 1820–1830.

[16] Andraos J, Sayed M. *J. Chem. Educ*. 2007, 84, 1004–1011.

[17] Andraos A, *The Algebra of Organic Synthesis: Design Strategy, Route Selection, and Optimization*, CRC Press-Taylor & Francis: Boca Raton, 2012.