Green Chemistry

RSCPublishing

SUPPLEMENTARY INFORMATION

Cite this: Green Chem., 2017, 19, 281

A deeper shade of green: inspiring sustainable drug manufacturing

Frank Roschangar, *^a Juan Colberg, ^b Peter J. Dunn, ^c Fabrice Gallou, ^d John D. Hayler, ^e Stefan G. Koenig, ^f David K. Leahy, ^g Michael E. Kopach, ^h Ingrid Mergelsberg, ⁱ John L. Tucker, ^j Roger A. Sheldon^{k,I} and Chris H. Senanayake^a

Received 19th October 2016, Accepted 15th November 2016

DOI: 10.1039/c6gco2901a

www.rsc.org/greenchem

Supplementary Discussion

Supplementary Discussion 1 | Achieving Consistency

The Green Aspiration Level (GAL) has been constructed on four pillars to ensure consistent application, namely (1) clearly defined synthesis starting points,¹ (2) unambiguous complete E factor (cEF)^{2,3} or Process Mass Intensity (PMI) waste metrics, (3) historical averages of industrial drug manufacturing waste, and (4) complexity of the drug's ideal manufacturing process (Supplementary Figure 6). cEF or PMI can be used interchangeably in GAL-based analysis enabling organizations using either to calculate their green performance scores. cEF and PMI differ by just one unit (Supplementary Equation 6) and share the same commercial waste goal for an average manufacturing step⁴ – the transformation-GAL or tGAL – that results in negligible numerical differences from the inclusion of one or the other. The pharmaceutical industry has generally adopted PMI. However, our publication utilizes cEF values due to literature prevalence and potentially broader appeal of E factors.⁵ It is important to note that all reaction and workup materials are included in the analysis, but excluded are reactor cleaning⁶ and solvent recycling.⁷

Standardized process starting points are a critical component of the GAL methodology. A starting material for some may be an intermediate for others. Until recently, the scientific community lacked an unambiguous definition of process starting points in the assessment of process greenness. This has been a bothersome source of inconsistency. Failure to define an appropriate starting material can lead to exclusion of significant amounts of intrinsic raw material waste created during earlier stages of manufacture. We therefore utilize these updated definitions of process analysis starting points to ensuring higher quality of data:⁸

 The material is commercially available from a major reputable chemical laboratory catalog company, and its price is listed in the (online) catalog. Materials requiring bulk or custom quotes do not qualify as process starting material.

AND

2) The laboratory catalog cost of the material at its largest offered quantity does not exceed US \$100/mol.

Therefore, published literature must be researched if the material does not qualify as process starting material in order to determine its correct intrinsic cEF. However, we realized that determination of literature cEF values is tedious and involves making assumptions since literature procedures are often incomplete compared to internal or external manufacturing batch records. Thus, standardizing Literature cEF quickly became a desirable goal. In order to facilitate literature analysis we introduced Supplementary Equation 7 that just requires determination of literature step count from ≤\$100/mol starting materials without having to retrieve literature waste information.⁹ The literature step multiplier of 37 kg/kg represents the average literature step cEF across the analyzed projects (Supplementary Table 1), so it equals their average literature cEF (76 kg/kg) divided by average literature step count (2.1). The process cEF and Relative Process Greenness (RPG) derived from the simplified calculated cEF literature values are shown next to their progenitors in Supplementary Table 3. We observe that average calculated and manually determined cEF and RPG values are comparable and within 10% of their means across the three development phases. Thus, we consider the simplified method sound and an important element to achieving consistency in green process analysis.

Supplementary Discussion 2 | Establishing Smart Green Manufacturing Goals

The GAL enables scientists and managers for the first time to define SMART (Specific, Measurable, Achievable, Resultsbased, Time-bound) and in particular Achievable green chemistry goals for any pharmaceutical manufacturing process while considering its complexity. Complexity is significant because drugs are not created equal, and their molecular and manufacturing process complexities vary greatly. Prior measures did not account for the complexity of a drug or the availability of technology to make it. While recent progress has measures been made to develop for molecular complexity,^{10,11,12,13} we selected the most facile process complexity measure based on Baran's ideality methodology (Supplementary Equation 8),¹⁴ which considers both molecular complexity and the degree of optimal implementation of available synthetic technology.¹⁵ For example, integration of this measure justly penalizes the use of protecting groups. Our approach is the most practical and allows any synthetic organic chemist to determine complexity of a given process within minutes. For example, we determine a complexity of 2 for the process for 'oxa acid', which is a synthetic precursor of Pradaxa. The process consists of 2 construction reactions and 2 concession steps involving a protecting group (Supplementary Figure 7).

We concede that our complexity measure is not impeccable since it does not give credit to innovation by way of reducing process step count. However, incorporation of Relative Complexity Improvement (RCI) into the Green Scorecard highlights the impact of innovation on overall process improvement. RCI specifically reflects improvements to process complexity through development of new synthetic approaches and methodologies (Supplementary Equation 5). The RCI concept is exemplified with an example from Lilly of a drug with representative complexity (Supplementary Figure 8).^{16,17} The progress made for processes C and B relative to A are indicated by RCI values of 38% and 25%, respectively,¹⁸ attributable to process innovation primarily driven by use of 2phenylsulfonyl pyridine 1 with concomitant elimination of all concession steps.

To determine the GAL goal as our overall process waste target, we multiply process complexity with tGAL. We simplify the original methodology that put forth phase-dependent goals and chose $tGAL = 26 \ kg/kg$ as the ubiquitous tGAL goal, independent of the drug's lifecycle status. 26 kg/kg is the average expected cEF-based process step waste per kg of commercial drug manufacture that was calculated from our data set by dividing average commercial cEF (156 kg/kg) by average commercial process complexity (5.9). The original value for the corresponding commercial cEF-based tGAL determined from the 2008 ACS GCI PR data was 19 kg/kg (ref. 15). The original tGAL values are shown in Supplementary Table 4. The inclusion of literature waste for non-commodity starting materials is the reason for the larger value of the updated tGAL. We note that the tGAL is considered a "moving target" as drugs become increasingly more complex, and therefore expect to periodically update tGAL. Conversely, as the industry more widely adopts green chemistry, we may expect tGAL to decrease with time.

Supplementary Discussion 3 | Input Fields and Output for the Green Scorecard Calculator

To use the Green Scorecard calculator, one can download the Excel file that is freely available from the IQ website (https://iqconsortium.org/initiatives/projects/green-aspiration-level), and then simply inputs project name and phase, complexity, sEF, and cEF for one or more completed campaign (input fields shaded in light green color, Supplementary Figure 9) in order to obtain the calculated RPG and green rating of the process. RPI, RCI, and PI are automatically calculated if two or more manufacturing campaigns are inputted. The first inputted campaign ought to be for an early development project to maintain consistency. The Green Scorecard graphic updates automatically and is ready for print and inclusion in project presentations to highlight the team's green process accomplishments.

Supplementary Equations

Supplementary Equation 1 | Determination of simple E Factor (sEF) a

sEF

$$=\frac{\sum m(Raw \ Materials) + \sum m(Reagents) - m(Product)}{m(Product)}$$

a, For reagents in solution, such as Grignard reagent solutions, only the mass component of the reagent will be considered as reagent mass. So 10 kg of a 10 weight% solution of a Grignard reagent in THF is counted as 1 kg of reagent and 9 kg of solvent.

Supplementary Equation 2 | Determination of Relative Process Greenness

$$RPG = \frac{GAL}{cEF} \times 100\%$$

Supplementary Equation 3 | Determination of overall Process Improvement (PI)

$$PI = \frac{RPI + RCI}{2}$$

Supplementary Equation 4 | Determination of Relative Process Improvement (RPI)

RPI

- = RPG(Current Process)
- RPG(First Development Process)

Supplementary Equation 5 | Measuring the Impact of Process Innovation via Relative Complexity Improvement (RCI)

$$RCI = 1 - \frac{Complexity (Current Process)}{Complexity (First Development Process)}$$

Supplementary Equation 6 | The complete E factor (cEF) and its simple mathematical Relationship to Process Mass Intensity (PMI)

 $cEF = PMI - 1 = \frac{\sum m(Raw \ Materials) + \sum m(Reagents) + \sum m(Solvents) + \sum m(Water) - m(Product)}{m(Product)}$

Supplementary Equation 7 | Simplifying Literature Waste Analysis $^{\rm a}$

$$cEF(literature) = no. of literature steps \times 37 \frac{kg}{kg}$$

a, Simplification applies to cEF but not sEF analysis.

Supplementary Equation 8 | Simple Definition of Complexity

Complexity = no.of construction steps + no.of strategic redox steps

Supplementary Figures

Supplementary Figure 1 | The Three Participants in a Sustainable Civilization



Supplementary Figure 2 | Boxplot of % Water and Solvent Use vs Project Phase



Supplementary Figure 3 | Graphic Definition of Literature Complexity as Component of External Manufacturing Complexity a



a, The place of manufacture, i.e. internal vs. external, is the defining element, irrespective of where the chemistry was developed. 1.



Supplementary Figure 4 | Commercial Pradaxa Process



Supplementary Figure 5 | Minitab's Probability Plot of Relative Process Greenness (RPG) for Early and Late Development and Commercial Processes

Supplementary Figure 6 | The Four Pillars of the Green Aspiration Level







Supplementary Figure 8 | Valuing Process Innovation for an Drug via RCI



Supplementary Figure 9 | Green Scorecard Calculator - Input and Output

INPUTS			tGAL =	26	kg/kg	Outputs				
Project Name	Pradaxa									
Analysis Date	19-May-16									
	Campaign #	Intended for <phase></phase>	Complexity	sEF	cEF	RPG	Rating	RPI	RCI	PI
First Dev Mfg								-	-	-
	2	Commercial	12	26.5	140.6	222%	Good			
Current	2	Commercial	12	26.5	140.6	222%	Good	-	-	-

Supplementary Table 1 | Green Analysis Data of 43 Small and 3 Large Molecule Drug Manufacturing Processes

Project Phase	Project #	Complexity	Internal Complexity	External Complexity	sEF	Internal cEF	External cEF	cEF	Lit Steps	Lit cEF ^a	calc cEF $^{\rm b}$	calc RPG ^c	GAL	RPG	% (Solvent + Water) ^d	% External Waste ^e	% Literature Waste ^f	External RPG ^g
Early De	evelopment																	
	1	2	0	2	8.1	0.0	208.2	208.2	4	37.5	318.7	16%	52	25%	96%	100%	18%	25%
	3	8	4	5	64.7	112.3	270.3	382.6	5	270.3	297.3	70%	208	54%	83%	71%	71%	48%
	4	13	11	2	59.9	308.8	112.9	421.7	3	112.9	419.8	81%	338	80%	86%	27%	27%	46%
	5	11	10	1	60.5	227.1	125.1	352.2	2	125.1	301.1	95%	286	81%	83%	36%	36%	21%
	6	16	6	10	86.0	350.0	489	839.0	1	18.6	857.4	49%	416	50%	90%	58%	2%	53%
	7	6	5	1	60.7	372.1	104.3	476.4	1	59.2	454.2	34%	156	33%	87%	22%	12%	25%
	9	0 10	5	5	67.5	362.7	516.3	2,746.0	4	167.2	2755.4	30%	208	30%	92%	42%	19%	25%
	10	10	7	10	89.0	355.9	390.4	746.3	1	27.6	755.7	58%	442	59%	88%	52%	4%	67%
	11	11	7	4	105.8	329.5	100.2	429.7	4	100.2	477.5	60%	286	67%	75%	23%	23%	104%
	12	8	-	-	84.7	-	-	244.3	3	122.9	232.4	90%	208	85%	65%	-	50%	-
	13	8	4	4	43.1	500.7	423.6	924.3	1	78.2	883.1	24%	208	23%	95%	46%	8%	25%
	14	8	3	5	14.1	101.6	52.5	154.1	5	52.5	286.6	73%	208	135%	91%	34%	34%	248%
	16	9	8	1	67.1	762.0	19.0	781.0	1	19.0	799.0	22%	230	30%	91%	2%	2%	137%
mean		9.4		3.6	90		304	793	2.6	113	777	47%	245	49%	87%	41%	21%	58%
median		8.5		2.0	67		208	611					221	41%	89%	36%	19%	31%
min		2		1	8		19	154					52	7%	65%	2%	2%	5%
max	م ماد الماد ال	17		10	271		1151	2,746					442	135%	96%	100%	71%	248%
Late De	velopment	3.7		2.9	67		285	655					96	33%	8%	24%	18%	62%
Late De	17	8	0	8	29.7	0.0	387.4	387.4	2	16.6	444.8	47%	208	54%	92%	100%	4%	54%
	18	18	7	11	113.5	290.7	281.5	572.2	13	281.5	771.7	61%	468	82%	80%	49%	49%	102%
	19	8	-	-	10.8	-	-	165.0	3	48.6	227.4	91%	208	126%	93%	-	29%	-
	20	4	-	-	6.7	-	-	115.3	1	18.1	134.2	77%	104	90%	94%	-	16%	-
	21	3	-	-	7.1	-	-	90.9	0	0.0	90.9	86%	78	86%	92%	-	0%	-
	22	4	2	- 5	8.2 47.1	181 7	297.5	479.2	5	297.5	366.7	39% 50%	104	39%	97%	- 62%	62%	44%
	24	13	3	10	64.9	70.4	316.7	387.1	3	135.3	362.8	93%	338	87%	83%	82%	35%	82%
	25	7	2	5	22.0	107.0	218.0	325.0	1	31.3	330.7	55%	182	56%	93%	67%	10%	60%
	26	6	2	4	32.4	355.6	104.0	459.6	1	55.2	441.4	35%	156	34%	93%	23%	12%	100%
	27	9	7	2	6.5	51.5	21.5	73.0	0	0.0	73.0	321%	234	321%	91%	29%	0%	177%
	28	10	1	9		80.0	79.0	159.0	2	31.0	202.0	129%	260	164%	-	50%	19%	296%
	29	0	4	7	77.6	50.4	601.6	652.0	0	162.6	525.4	190%	260	190%	-	54%	25%	211%
	31	5	1	4	14.5	139.0	210.0	349.0	0	0.0	349.0	37%	130	32%	96%	52% 60%	23%	50%
average	-	8.0		6.5	34		236	308	2.1	72	315	90%	208	96%	91%	61%	17%	109%
median		8.0		6.0	22		218	325					208	82%	92%	60%	12%	82%
min		3		2	7		22	73					78	32%	80%	23%	0%	30%
max		18		11	114		602	652					468	321%	97%	100%	62%	296%
standar	d deviation	3.7		2.7	32		161	177					96	76%	5%	23%	19%	80%
Comme	32	12	4	8	26.5	70.8	69.8	140.6	1	7.7	169.9	184%	312	222%	81%	50%	5%	298%
	33	11	7	4	9.9	50.4	35.1	85.5	5	35.1	235.4	121%	286	335%	88%	41%	41%	296%
	34	7	7	0	13.0	153.0	0	153.0	0	0.0	153.0	119%	182	119%	92%	0%	0%	-
1	35	3	-	-	7.6	-	-	64.5	0	0.0	64.5	121%	78	121%	88%	-	0%	-
1	36	3	-	-	8.6	-	-	37.2	0	0.0	37.2	210%	78	210%	77%	-	0%	-
1	37	4	-	-	3.7	-	-	53.5	1	18.1	72.4	144%	104	194%	93%	-	34%	-
1	39	4	- 3	- 2	9.7	112 5	84.2	196.7	2	0.0 84 2	186 5	4/%	104	4/%	90% 95%	43%	U% 43%	62%
1	40	7	2	5	22.0	76.3	175.7	252.0	3	123.0	240.0	76%	182	72%	91%	70%	49%	74%
1	41	5	2	3	13.8	65.0	79.1	144.1	3	79.1	176.0	74%	130	90%	90%	55%	55%	99%
1	42	6	2	4	27.7	126.0	71	197.0	0	0.0	197.0	79%	156	79%	86%	36%	0%	146%
	43	4	1	3	25.8	139.0	184	323.0	1	53.0	307.0	34%	104	32%	92%	57%	16%	42%
average		5.9		3.6	15		87	156	1.3	33	172	106%	154	132%	89%	44%	20%	145%
mic		5.0		3.5	12		/5	149	IGAL (updated	u) = 26			130	105%	91%	46%	11%	99%
max		3 12		8	28		184	3/323					312	32%	96%	70%	0%	42%
standar	d deviation	2.8		2.2	8		59	83					73	86%	5%	19%	21%	101%
OVERA	LL AVERAGE							446	2.1	76		78%			89%	48%	20%	93%
OVERA	LL COUNT (N)								43	43					41	34	43	33
									Lit Step cEF	= 37								
Others	(Polypeptides,	Bioconjugate	es, NBEs. etc)	- examples w	here GAL	is not appli	cable		-									
1	44	10	-	-	238.0	-	-	6,457.0	5	2102.8	4539.2	6%	260	4%	96%	-	33%	-
1	46	7	0	7	130.6	0.0	2472	2,472.0	0	0.0	2472.0	7%	182	20%	95%	100%	0%	5%

a, Lit cEF = cEF derived from researching literature procedures; b, calc cEF = cEF – Lit cEF + Lit Steps *x* Lit Step cEF = cEF – Lit cEF + Lit Steps *x* 37 kg/kg; c, calc RPG = Complexity *x* tGAL / calc cEF = Complexity *x* 26 kg/kg / calc cEF; d, % (Solvent + Water) = (cEF - sEF) / cEF; e, % External Waste = External cEF / cEF; f, % Literature Waste = Lit cEF / cEF; g, External RPG = External Complexity *x* tGAL / External cEF.

Supplementary Table 2 | Current and Initial Drug Manufacturing Waste Data ^a

Phase of Drug	2016 cEF Mean	2008 cEF Mean (ACS GCI PR) [kg/kg]
	$[Kg/Kg]$ (N \sim)	(N)
Early Development	793 (16)	553 ° (11)
Late Development	325 (14)	254 ^d (33)
Commercial	156 (12)	152 (7)

a, the ACS GCI PR presentation of ref. 25 in the main article provides the median PMIs, so the original dataset was reanalyzed, and the cEF means were determined to allow comparison to the 2016 data; b, N = number of analyzed drug manufacturing processes; c, weighted mean of preclinical and Phase 1 drugs; d, weighted mean of Phase 2 and Phase 3 drugs.

Supplementary Table 3 | Assessing the Impact of Literature-derived Waste ^a

Phase of Drug	N ^b	cEF	RPG	calc cEF ^c	calc RPG ^c
		[kg/kg]	[kg/kg]	[kg/kg]	[kg/kg]
Early Development	16	793	49%	777	47%
Late Development	14	325	80%	332	74%
Commercial	12	156	132%	172	106%

a, all figures represent the means; b, N = number of manufacturing process data sets; c, "calc" reflects incorporation of the calculated literature cEF based on Supplementary Equation 3 – see Supplementary Table 1.

Supplementary Table 4 | Originally Reported Goals for Process Step Waste

Phase of Drug	sEF tGAL	cEF tGAL			
	[kg/kg]	[kg/kg]			
≤ PoC ^a	5	34			
> PoC	3	19			

a, PoC = Proof of Concept, refers to exploratory (non-pivotal) Phase IIa studies in patients or healthy volunteers that evaluates clinical efficacy, Pharmacodynamics, or biological activity as primary endpoint.

Supplementary References

- The \$100 per mol laboratory catalog pricing requirement described in Supplementary Discussion 1 does not apply to reagents, catalysts, ligands, and solvents, since they are produced for widespread application and are not specific to the process being evaluated.
- Since the original E factor has been applied inconsistently, the cEF metric was introduced for the purpose of GAL analysis. cEF accounts for all process reaction and process workup materials, including raw materials, intermediates, reagents, process aids, solvents, and water.
- 3. All E factors reported herein represent the cEF or sEF contributions of the overall manufacturing process or the sub-process (e.g. external cEF, literature cEF) to produce 1 kg of drug substance.
- 4. We define a step as a chemical operation involving one or more chemical transformations that form and/or break covalent or ionic bonds and lead to a stable and isolable intermediate, but not necessarily include its isolation. Examples: Simultaneous removal of two or more protection groups involves multiple transformations, yet it is carried out in one chemical operation → counted as one step Sequential transformations via a stable and isolable intermediate that are carried out in two operations but

without intermediate workup \rightarrow counted as two steps • Formation of covalent bonds or salts that occur during workup \rightarrow not counted as an extra step • Separate operation of salt formation from an isolated intermediate \rightarrow counted as one step • Isolation of a product, following work-up, as a solution that can be stored \rightarrow counted as one step.

- 5. A SciFinder search for the terms 'Process Mass Intensity', and 'E factor' and 'Environmental impact factor' on Nov. 14, 2016 revealed that the PMI concept was present in 12, 8, 9, and 12 publications for the years 2013-2016, respectively, while the E factor concept was mentioned 39, 45, 57, and 46 times (76-86%), respectively.
- 6. The GAL considers only direct process materials, i.e. materials used in the chemical steps and their workups. It does not include solvents and aqueous detergents required for reactor and equipment cleaning between batches or steps, nor the frequency and duration of the equipment and facility specific cleaning operations. These parameters are considered for comprehensive environmental impact in Life Cycle Assessment (LCA) analysis.
- In US pharmaceutical manufacturing, recycling accounts for 25% of waste handling, while energy recovery burning and treatment constitute 38% and 35%, based on 2012 data from 'The Right-To-

- 8. The \$100 per mol commodity pricing criterion was established in ref. 15 of the main article based on the author's professional experience. The authors of this manuscript consider this figure appropriate and helpful for providing a consistent analysis.
- 9. If a detailed procedure is available for a particular literature step, its calculated waste can be used in place of the 37 kg/kg default value.
- J. Li and M. D. Eastgate, Current Complexity: a Tool for Assessing the Complexity of Organic Molecules. *Org. Biomol. Chem.* 2015, 13, 7164–7176.
- D. P. Kjell, I. A. Watson, C. N. Wolfe and J. T. Spitler, Complexity-Based Metric for Process Mass Intensity in the Pharmaceutical Industry. *Org. Process Res. Dev.* 2013, **17**, 169– 174.
- R. P. Sheridan, *et al.*, Modeling a Crowdsourcing Definition of Molecular Complexity. J. Chem. Inf. Model. 2014, 54, 1604–1616.
- M. F. Faul, *et al.*, Part 2: Designation and Justification of API Starting Materials: Current Practices across Member Companies of the IQ Consortium. *Org. Process Res. Dev.* 2014, 18, 594–600.
- 14. Besides offering simplicity, the GAL's process complexity model was selected vs. the alternative structural complexity measures due to its inherent ideality-derived consideration for available synthetic methodology.
- 15. See main article ref. 16: it defines Construction Reactions (CR) as chemical transformations that form skeletal C-C or C-heteroatom bonds. Strategic Redox Reactions (SRR) are construction reactions that directly establish the correct functionality found in the final product, and include asymmetric reductions or oxidations. All other types of non-strategic reactions are considered as Concession Steps (CS), and include functional group interconversions, non-strategic redox reactions, and protecting group manipulations.
- M. E. Kopach, et al., Process Development and Pilot-Plant Synthesis of (2-Chlorophenyl)[2-(phenylsulfonyl)pyridin-3yl]methanone. Org. Process Res. Dev. 2010, 14, 1229–1238.
- M. E. Kopach, M. M. Murray, T. M. Braden, M. E. Kobierski, O. L. Williams, Improved Synthesis of 1-(Azidomethyl)-3,5-bis-(trifluoromethyl)benzene: Development of Batch and Microflow Azide Processes. Org. Process Res. Dev. 2009, 13, 152–160.
- 18. RCI (Process B) = $1 (\frac{6}{2}) = 0.25$. RCI (Process C) = $1 (\frac{5}{2}) = 0.38$