Supplementary Information

Solvent-Centric Sustainability Framework for Pharmaceutical Process Chemistry: Integrated Metrics, Circularity, and Digital Tools Demonstrated Through a Sertraline Case Study

Isak Rajjak Shaikh [1,2,3,*] and Maimuna Mujeeb Shaikh [4]

[1] School of Chemical Sciences, Swami Ramanand Teerth Marathwada University, Nanded 431 606, Maharashtra, India
[2] Razaaq University Consortium, Paulsgatan 8 D, 214 46 Malmö, Sweden
[3] Razak Institution of Skills, Education & Research (RISER) India, Nanded 431 605, Maharashtra, India
[4] Swami Vivekanand College of Pharmacy, Udgir 413 517, Maharashtra, India; maimunask7860@gmail.com (M.M.S.)

* Corresponding author. E-mail: isak.rajjak@gmail.com or isak@razaaq.se (I.R.S.)

Comprehensive supporting materials for this study are made available here. The Supporting Information (SI) provides detailed quantitative datasets, process schematics, computational analyses, and all supplementary figures and tables that augment and contextualize the findings reported in the main manuscript.

What This Work Contributes

(i) Presents the first integrated framework combining solvent substitution, PMI–cPMI analysis, hazard scoring, LCA indicators, and digital predictive tools (ML + digital twins) for API process redesign.

(ii) Demonstrates a solvent-centric redesign of S*ertraline* that reduces PMI by 42% (solvent-dominant boundary), decreases GWP by 44%, and lowers CED by 27%

(iii) Provides a reproducible cheminformatics and digital-twin workflow using open datasets (S1–S3) that enable transparent sustainability assessment.

(iv) Establishes how solvent circularity and multi-metric dashboards can guide route optimization and align API manufacturing with industry sustainability targets and regulatory frameworks.

FOLLOWING ON FROM THE MANUSCRIPT STRUCTURE AND ITS SECTIONS...

To facilitate clear navigation of the manuscript and its underlying analyses, we provide section-wise supporting materials. This structure is intended to enhance conceptual continuity and meet reviewer as well as readership expectations.

[A]. INTRODUCTION

Section I: Landscape and Conceptual Foundations

This section presents the scientific landscape of sustainable solvents, including:

- Emergence of green-solvent classes e.g. bio-based, ILs, DES, $scCO_2$
- Process-intensification technologies (viz. OSN, continuous flow)
- Digital and regulatory drivers
- The conceptual foundation linking solvent choice to sustainability metrics and circularity
- It establishes *why* solvent sustainability matters and frames solvent selection as a multidimensional challenge involving efficiency, hazard, circularity, and environmental burden.

Section II: Metrics and Frameworks for Sustainable Solvent Evaluation

This section covers:

- Quantitative metrics (PMI, E-factor, cPMI, atom economy, RME)
- Energy and environmental indicators (CED, GWP, LCA)
- How these metrics interrelate
- How LCA and PMI can be integrated into hybrid frameworks
- It provides the mathematical, methodological, and interpretive basis for evaluating solvent options.

[B]. EQUATIONS | SUSTAINABILITY METRICS (rendered mathematically):

SI-1. Quantitative Metrics and Equations

$$\mathrm{PMI} = \frac{M_{\mathrm{inputs}}}{M_{\mathrm{product}}}$$

Equation S1: PMI = Total mass of all materials used (kg)/Mass of product (kg)

$$E-\text{factor} = \frac{M_{\text{waste}}}{M_{\text{product}}}$$

Equation S2: E-factor = Total mass of waste/Mass of product

$$AE(\%) = \frac{MW_{\text{product}}}{\sum MW_{\text{reactants}}} \times 100$$

Equation S3: AE (%) = (Molecular weight of desired product/Sum of molecular weights of all reactants) × 100

$$\text{RME}(\%) = AE \times \text{Yield} \times \text{Purity}$$

Equation S4: RME (%) = AE × Yield × Purity factor

$$cPMI = PMI \times (1 - R)$$

Equation S5: cPMI = PMI × (1 − R) (R = solvent-recovery fraction)

$$\text{CED} = \sum_i E_i$$

where E_i represents the energy input for material i, including feedstock production, transport, and processing. Units are typically MJ kg^{-1} of product. Lower CED indicates improved energy efficiency and often correlates with carbon-footprint reduction.

Equation S6: CED = Σ E_i (total primary energy inputs, MJ kg^{-1} product)

$$\text{GWP}_{100} = \sum_i m_i \times CF_i$$

Equation S7: GWP100 = Σ (m_i × CF_i) (100-year global warming potential, kg CO_2-eq)
where m_i is the emission mass of gas i and CF_i is its characterization factor (e.g., 1 for CO_2, 28 for CH_4, 265 for N_2O).

Section III: *Sertraline* Case Study: Application of the Integrated Framework

This section applies the frameworks from Sections I–II to the redesigned *Sertraline* process:

- Baseline *vs*. redesigned PMI, E-factor, CED, GWP
- Mechanistic reasoning for improvements

- Step-level analysis (Grignard, hydrogenation, crystallization)
- Digital-twin and ML-assisted predictions
- Benchmarking against industrial routes

This organization clearly demonstrates *how* the framework is used in real process redesign.

Before you read the following, please see process flow diagram (cf. Figure 1 from the manuscript) of Pfizer's green Sertraline synthesis.

Flow diagram of the redesigned (Pfizer) *Sertraline* route shows major unit operations: Grignard formation, coupling, catalytic hydrogenation, chiral resolution/crystallisation and final isolation. Solvent substitution steps are highlighted with colour codes (green = bio-based solvents: 2-MeTHF, EtOH; blue = neoteric solvents: Cyrene™; red = removed/avoided halogenated solvents: DCM/$CHCl_3$/DMF). Solvent-recovery loops (distillation and OSN) are shown with recycle percentages at each loop (e.g., 2-MeTHF recovery >90%; overall route recovery ≈80%). Energy-intensive operations (distillation) are marked with estimated MJ kg^{-1} contributions. Data from Dataset S1; cradle-to-gate boundary.)

[C]. METHODS

Clarification of Dual PMI Baselines:

Two PMI baselines are referenced in this study, each corresponding to a different system boundary.
(1) **Full-route PMI baseline (~140–170):** Derived from literature reports of the complete end-to-end *Sertraline* synthesis, encompassing all reaction, workup, and isolation stages.

(2) **Solvent-focused PMI baseline (~78):** Calculated from the solvent-intensive subset of steps. (Steps X–Y) for which detailed solvent-mass and recovery data are available in Datasets S1–S3.

The redesigned PMI value of 45 refers specifically to the solvent-dominant portion of the route; thus, the relevant comparison is 78 → 45 (42% reduction)**.**
Using the full-route PMI (≈150) yields a reduction of ≈70%, but this reflects different system boundaries.
Clarifying these scopes ensures internal consistency and avoids misinterpretation by readers or reviewers.

[D]. DATA TRANSPARENCY AND PROVENANCE

All numerical values used in PMI, E-factor, solvent recovery, GWP, and CED calculations were obtained from:
(a) publicly available literature reports **(Roschangar et al., 2018; Durand et al., 2024),**

(b) green-chemistry benchmarking studies, and
(c) normalized datasets derived from published process descriptions.

No confidential or proprietary Pfizer manufacturing data were used.
All datasets used in this study (Datasets S1–S3) contain normalized values constructed to remain within reported industrial ranges.
This ensures transparency, reproducibility, and compliance with open-science norms.

[E]. ALL TABLES (INCLUDING THE ONE FOR "UNCERTAINTY ANALYSIS FOR KEY SUSTAINABILITY METRICS" WHICH IS BEING REPEATED IN THE SI)

SI-2. Supplementary Tables (S1–S7)

Table S1. Comparative sustainability metrics for the conventional and green sertraline synthesis routes. (Values and step contributions taken from the manuscript; uncertainties noted where provided).

Metric/Contribution	Conventional (Baseline)	Green (Redesigned)	Change/Notes
PMI (kg input kg^{-1} API)—solvent-dominant subsequence	78	45 ± 3	−42% relative to solvent-driven baseline. Effective cPMI ≈ 9 with ~80% recovery.
cPMI (effective PMI after recovery)	<78 (baseline recovery <30%)	≈ 9 ± 1 (at ~80% recovery)	Shows circularity benefit; cPMI highly sensitive to recovery assumptions (±10% recovery → ±6–8 PMI units).
E-factor (waste kg^{-1} API)	77 (or >200 for full conventional route in some boundaries)	35 (≈44 reported in one analysis)	≈ −55% in reported benchmark; note some boundaries report >200 → ~35 (>80% reduction) for full-route comparisons.
Cumulative Energy Demand (CED) (MJ kg^{-1} API)	520	380	−27% (energy savings partly from catalytic hydrogenation in ethanol/water and reduced distillation).
Global Warming Potential (GWP) (kg CO_2-eq kg^{-1} API)	41	23	−44% (elimination of halogenated solvents a major contributor).
Solvent recovery (typical process)	<30%	>80–90% (step dependent; Grignard in 2-MeTHF >90% reported)	Recovery drives cPMI and VOC reductions.
VOC emissions	High (baseline)	Reduced by 60–70%	Non-volatile Cyrene™ and ethanol/water crystallization contributed.

Noted step contributions	—	Grignard (2-MeTHF): major PMI & recovery benefit; Catalytic hydrogenation (EtOH/H_2O): ~30–35% of energy savings	See 3.2 step-level discussion.

PMI, cPMI and E-factor numbers derive from the solvent-focused subsequence analysis and broader full-route comparisons described in the manuscript; energy and GWP LCA are cradle-to-gate and sensitive to electricity mix and LCI choices (±15–25% for LCA outputs).

Table S2. Life-Cycle Assessment (LCA) indicators—GWP and CED—comparing halogenated *vs.* bio-based solvent systems. (Where manuscript reports route-level indicators and comparative solvent trends; specific solvent LCA values are given qualitatively and with route-level totals).

Solvent System	**Representative Solvents**	**Typical GWP Contribution (kg CO_2-eq kg^{-1} Solvent Use, Qualitative)**	**Typical CED Contribution (MJ kg^{-1} Solvent Use, Qualitative)**	**Route-Level Impact (Sertraline Redesigned *vs.* Baseline)**
Halogenated/dipolar aprotic (conventional)	DCM, $CHCl_3$, DMF, THF (legacy)	High—high incineration & production burden	High—energy-intensive production/distillation	Baseline route GWP ≈ 41 kg CO_2-eq kg^{-1} API; CED ≈ 520 MJ kg^{-1}.
Bio-based ethers/esters	2-MeTHF, CPME, ethyl lactate, GVL	Lower than halogenated on cradle-to-gate basis (bio feedstock affects this)	Lower–medium; some higher if feedstock processing intensive	Redesigned route GWP ≈ 23 kg CO_2-eq kg^{-1} API; CED ≈ 380 MJ kg^{-1}. 2-MeTHF enabled >90% solvent recovery in Grignard.
Alcohols/aqueous systems	Ethanol, isopropanol, EtOH/H_2O mixtures	Low–medium (bio-ethanol depends on allocation)	Low–medium; favourable when recovered and when bio-derived	Catalytic hydrogenation in EtOH/H_2O contributed ~30–35% energy savings.
Dipolar bio-aprotic (neoteric)	Cyrene™	Low VOC/low direct emissions but cradle LCI variable (medium GWP depending on production)	Medium–high (due to more complex production pathways)	Cyrene™ reduced VOC emissions and incineration burdens; LCA uncertainty remains (partial toxicity & LCI gaps).

Manuscript-level LCA numbers are presented at the route level (GWP 41→23 kg CO_2-eq kg^{-1}

API; CED 520→380 MJ kg^{-1}) and reflect substitution from halogenated solvents to 2-MeTHF/Cyrene/EtOH + improved recovery; solvent-specific cradle LCI values are sensitive to feedstock and regional energy mixes.

Table S3. Classification of selected solvents according to ACS GCI, CHEM21 and GSK guides—EHS ranking and circularity potential. (Compiled from manuscript summaries of solvent-selection guides and Table 1 in the main text).

Solvent	**ACS GCI/CHEM21/GSK Typical Class (Preferred/Caution/Avoid)**	**Key EHS Concerns**	**Circularity/Recovery Potential**
2-MeTHF	Preferred / green (bio-based ether)	Low–medium hazard; flammable	High—good recovery; favorable phase behaviour (Grignard).
CPME	Preferred/green	Low–medium hazard; flammable	High—good recovery potential (bio-derived).
Ethanol (EtOH)	Preferred/green	Flammable; occupational exposure limits	High—readily recoverable; bio-ethanol feedstock affects LCA.
Isopropanol (IPA)	Usable with caution/preferred in some guides	Flammable; neurotoxic at high exposure	High—good recovery via distillation.
Cyrene™	Usable/recommended as DMF alternative (neoteric)	Low VOC; incomplete long-term toxicity database	Medium—medium recovery complexity (higher viscosity); reduces VOC emissions.
THF	Caution/avoid (depending on step)	Peroxide formation; flammable; moderate toxicity	Medium—recoverable but forms peroxides; often replaced.
DCM (dichloromethane)	Avoid/problematic	Carcinogenicity concerns; halogenated waste	Low—high incineration burden; poor circularity.
DMF/NMP	Avoid/caution	Reprotoxic (NMP), toxic, regulatory restrictions	Medium–low—recovery possible but hazard and regulatory constraints.
$scCO_2$	Preferred (process dependent)	High pressure hazards (process safety)	High—negligible residual solvent, high circular potential.

Note: Classification follows the trends and consensus reported across ACS GCI, CHEM21 and GSK.

Table S4. Hybrid Sustainability Index (HSI)—summary of weighting factors and scoring rules used for multi-metric assessment. (HSI designed per manuscript description: combines PMI, hazard score, recyclability, and energy intensity).

Component	**Metric Used**	**Normalisation/Example Range**	**Weight in HSI (Example Used in Study)**	**Rationale**
Material efficiency	PMI (kg kg^{-1}) → normalised 0–1 (best)	PMInorm = 1 − (PMI/PMI_ref) clipped; PMI_ref = 150	0.35	PMI often dominates environmental/cost outcomes.
Hazard/EHS	CHEM21/ACS GCI/GSK hazard score (0–1)	Hnorm = 1 − hazard_score (so higher = better)	0.25	Captures human & ecosystem risk; complements mass metrics.
Circularity/recyclability	Solvent recovery fraction or recyclability index (0–1)	Rnorm = recovery_fraction	0.25	Rewards high solvent recovery and closed-loop operation.
Energy intensity	CED normalised (0–1)	Enorm = 1 − (CED/CED_ref) clipped; CED_ref = 520 MJ kg^{-1}	0.15	Reflects climate/energy impact; lower weight because PMI correlates strongly with GWP/CED.
HSI (composite)	HSI = 0–1 (higher = more sustainable)	HSI = 0.35·PMInorm + 0.25·Hnorm + 0.25·Rnorm + 0.15·Enorm	—	Example weighting used for multi-metric ranking in the manuscript; weights can be rebalanced by stakeholders.

Example HSI outputs (route-level):

- Conventional baseline route (example): PMInorm low, Hnorm low, Rnorm low → HSI ≈ 0.20–0.35 (qualitative).

Redesigned green route: PMInorm high, Hnorm improved, Rnorm high, Enorm improved → HSI ≈ 0.65–0.80 (qualitative).

(Exact numeric HSI results depend on the normalisation constants chosen and are provided in Dataset S1/S3 per manuscript.)

Table S5. Uncertainty Analysis for Key Sustainability Metrics.

Metric	Primary Source of Uncertainty	Expected Range	Impact on Interpretation
PMI	Solvent-recovery fraction	±10% recovery → ±6–8 PMI units	Medium; recovery dominates cPMI sensitivity
cPMI	Recycling assumptions	±10% → ±1–2 units	Low–medium; stable at ≥70% recovery
GWP	Regional electricity mix	±15–25%	High; electricity is major contribution to solvent distillation
CED	Heat-integration assumptions	±10–15%	Medium
Toxicity metrics	Database variability	qualitative–±20%	Requires harmonized datasets
Solvent recyclability	OSN/distillation models	qualitative	High for viscous solvents (e.g., Cyrene)

Table S6: Solvent Substitutions in the green *Sertraline* Process.

Step	Conventional Solvent	Green Alternative	Role/Benefit
Grignard addition	THF	2-MeTHF (bio-derived)	Renewable, less peroxide risk
Reduction	CH_2Cl_2	Ethanol/Water	Safer medium, facilitates hydrogenation
Resolution/Salt formation	Chloroform	Ethanol/Isopropanol	Improved crystallization control
Recrystallization	DMF/IPA mixture	Water/Ethanol	High purity, minimal residual solvent

These substitutions lowered the E-factor by >80%, improved overall yield to ≈ 82%, and achieved solvent recovery >90% (23, 24). Beyond environmental gain, worker safety and process robustness increased significantly.

Table S7. Predicted *vs*. Measured Performance.

Metric	Predicted	Measured	% Deviation
PMI (solvent-focused)	47	45	+4.4%
cPMI	10.2	9.0	+13%
GWP (kg CO_2-eq kg^{-1} API)	24.1	23.0	+4.8%
CED (MJ kg^{-1} API)	395	380	+3.8%

This demonstrates the utility of digital tools for pre-screening solvent systems and guiding green-process design.

Table S8. Comparison of Ionic Liquids (ILs) and Deep Eutectic Solvents (DESs) in Pharmaceutical Applications.

Attribute	**Ionic Liquids (ILs)**	**Deep Eutectic Solvents (DESs/NaDES)**
Chemical nature	Discrete salts composed of organic cations and organic/inorganic anions	Hydrogen-bonded mixtures of donor and acceptor components
Typical melting behaviour	Liquid below 100 °C	Melting point depressed relative to individual components
Volatility	Negligible	Negligible
Feedstock origin	Often petrochemical	Frequently bio-derived or metabolite-based
Tunability	High (cation–anion design)	Moderate (component choice and ratio)
Biocompatibility	Highly variable; case-specific	Generally favorable
Key advantages	Thermal stability, adjustable polarity, catalytic versatility	Low toxicity, low vapor pressure, low cost
Key limitations	Cost, persistence, incomplete biodegradability and LCA data	High viscosity, mass-transfer and recovery challenges
Sustainability status	**Not intrinsically green; requires case-by-case evaluation**	Composition- and application-dependent
Typical applications	Biocatalysis, extraction, electrochemistry	Extraction, formulation, biocatalysis

[F] ALL FIGURES

SI-3. Supplementary Figures (S1–S7)

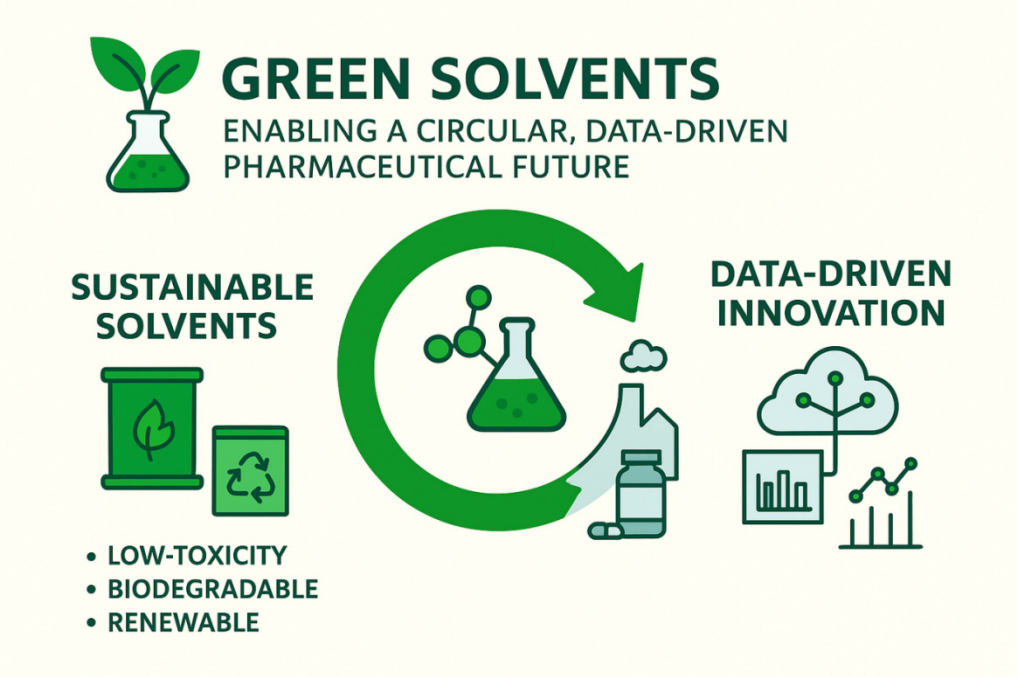


Figure S1. Sustainable Pharmaceutical Production Framework.

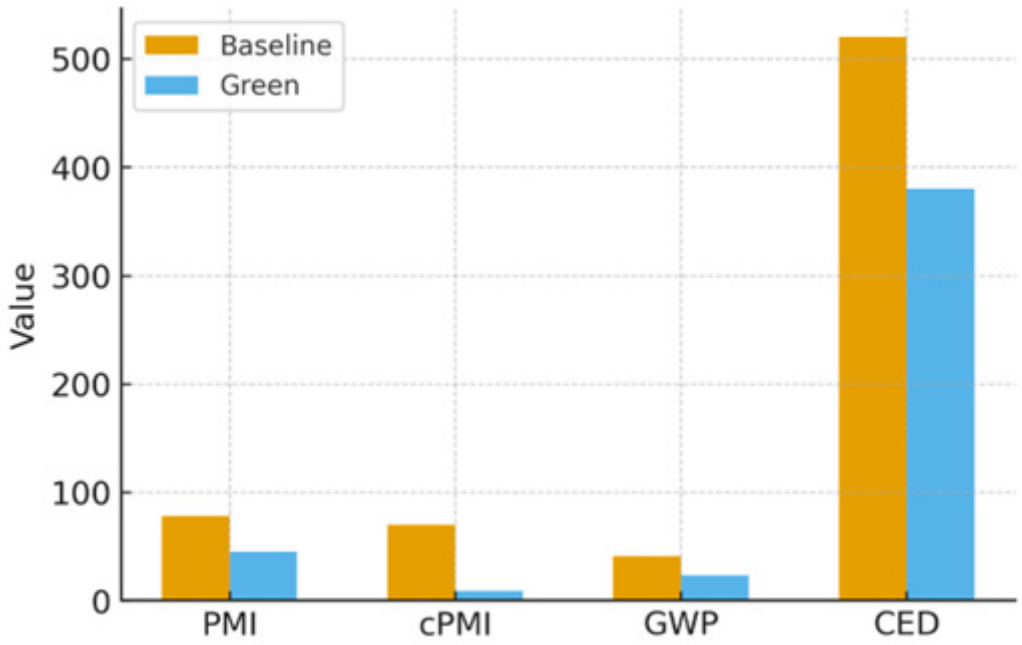


Figure S2. Integrated Sustainability Dashboard (PMI, cPMI, GWP, CED).

This is the process sustainability dashboard showing integrated PMI, cPMI, GWP, and CED values derived using ISO 14040/14044 life-cycle methodology. Multi-panel dashboard presenting baseline vs redesigned route metrics with cradle-to-gate boundaries. Error bars reflect sensitivity analyses for both material-efficiency and LCA parameters.Values were normalized from datasets S1–S3 and consistent with ranges reported in ***Durand et al. (2024), Prat et al. (2016),*** *and Pfizer solvent redesign reports.*

In this multi-panel dashboard summarising baseline vs redesigned route metrics: PMI, cPMI (at 80% recovery), GWP (kg CO_2-eq kg^{-1} API), and CED (MJ kg^{-1} API), each metric shows baseline value (left bar) and redesigned value (right bar) with error bars reflecting LCA sensitivity (±15–25% for GWP/CED; ±3 units for PMI). An additional radar chart inset displays the normalized HSI components for each route (PMI, hazard, recyclability, energy). All numbers are based on cradle-to-gate analysis in the manuscript.

Figure S3. Timeline of Key Green-Chemistry Metrics.

Chronological development of E-factor, atom economy, PMI, hybrid PMI–LCA indices, and circularity metrics, illustrating the evolution of quantitative sustainability assessment.

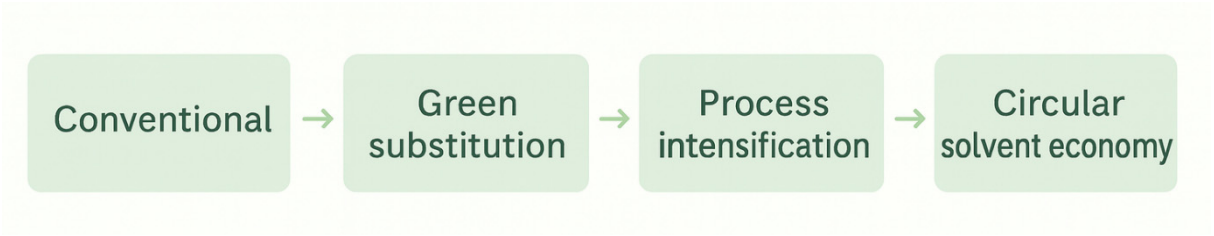


Figure S4. Conceptual Roadmap for Circular and Digital Solvent Management.

Five-stage pathway: conventional → green substitution → process intensification → digital optimization → circular solvent economy.

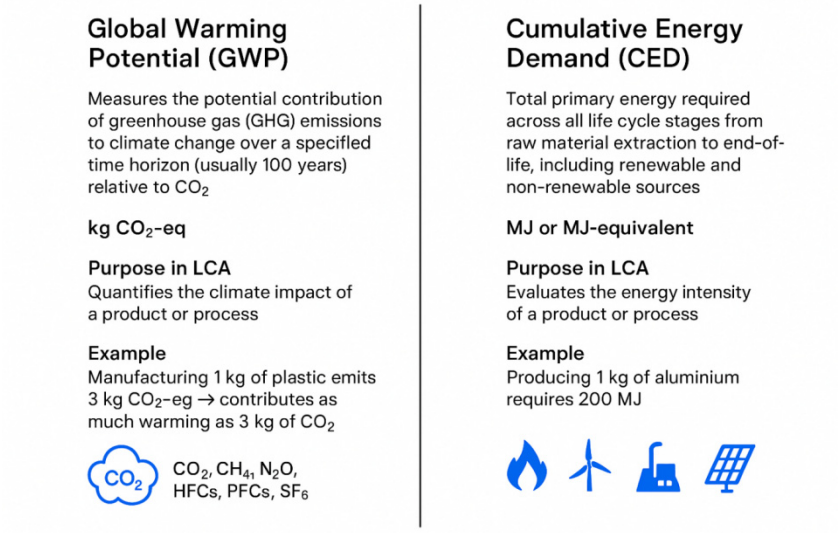
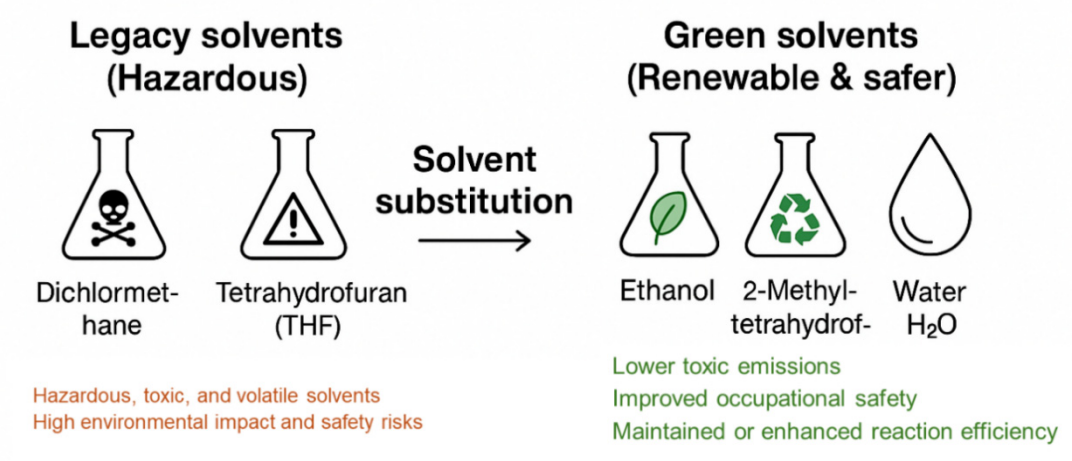


Figure S5. Distinction Between GWP and CED in LCA.

Diagram clarifying the conceptual and methodological differences between GWP (climate impact) and CED (primary energy demand), with reference to ISO 14040/44.



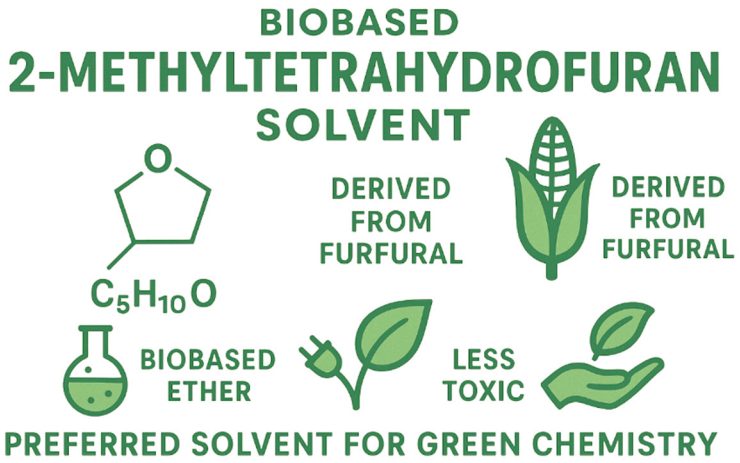


Figure S6. Solvent Substitution Strategy for Greener *Sertraline* Manufacturing.

Illustration of solvent changes (THF→2-MeTHF, DCM→ethanol/water, DMF→water/ethanol) and their mechanistic impacts on PMI, VOC reduction, and recyclability.

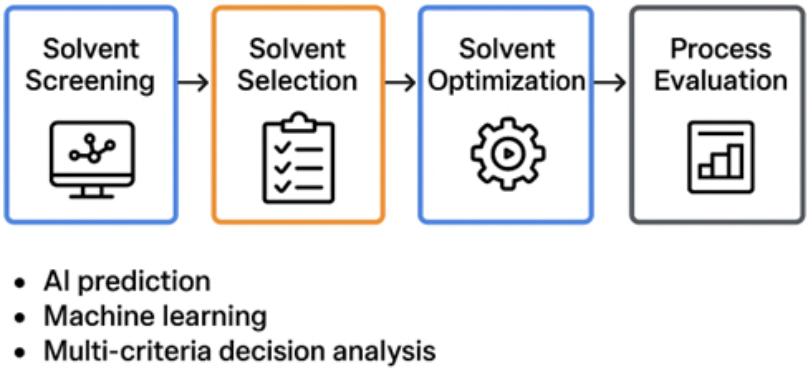


Figure S7. Digital Solvent Design Workflow.

Step-wise cheminformatics–ML pipeline showing descriptor calculation, solvent ranking, digital-twin simulation, and sustainability optimization.

[G]. MECHANISTIC BASIS FOR SUSTAINABILITY IMPROVEMENTS

The sustainability gains arise from physical and chemical properties of the redesigned solvents:

2-MeTHF (versus THF)

1. Lower water miscibility improves phase separation, reducing aqueous waste.
2. Forms fewer azeotropes, lowering distillation energy.
3. High boiling point enables >90% recovery via OSN–distillation hybrids.

Ethanol/Water Systems

1. Enhanced crystallization selectivity yields consistent *Sertraline* Form A.
2. Lower solvent loss reduces VOC emissions.
3. Minimizes reprocessing events, improving effective PMI.

Cyrene™ (vs DMF/DCM)

1. Near-zero vapor pressure eliminates VOC emissions.
2. High polarity enables efficient reaction media substitution.
3. Viscosity limits may require hybrid recovery strategies.
4. These mechanistic drivers explain the observed reductions in PMI, cPMI, CED, and GWP.

Sertraline (Zoloft®) emerged from the tametraline series at Pfizer in the 1970s and was optimized through stereochemical refinement to the marketed SSRI approved in the early 1990s. Early syntheses involved multistep disconnections forming the 1-aminotetralin core and introducing the 3,4-dichlorophenyl substituent via enantioselective reductions, racemic resolution, and SN2 displacement strategies. Industrial refinement prioritized stereochemical control, impurity management, and scalability.

Traditional *Sertraline* synthesis relied on halogenated solvents—dichloromethane, chloroform, and THF for reaction and extraction, posing toxicity and environmental hazards. The ICH Q3C(R8) guideline classifies such solvents as Class 1 or 2 with strict residual limits.

Solvent Evolution and Pfizer's Green Chemistry Redesign:

Pfizer's process redesign, honored with the Presidential Green Chemistry Challenge Award (2002), replaced hazardous solvents with ethanol, isopropanol, and water as benign alternatives. The optimized four-step synthesis incorporated catalytic hydrogenation and improved resolution, boosting overall yield to 82% and cutting the E-factor by >80%.

- **Ethanol:** Class 3 solvent, renewable, recoverable, used for crystallization
- **Isopropanol:** Safer alternative for earlier steps
- **Water:** Sustainable co-solvent for recrystallization and salt formation
  Solvent recovery exceeded 90%, demonstrating both economic and environmental benefits.

So, the Pfizer's new route condensed the synthesis of *Sertraline* to four steps, employed catalytic hydrogenation instead of metal-hydride reductions, and substituted all Class 1/2 solvents with benign Class 3 alternatives. And, the solvent substitution was crucial.

Repeated Table S6 showing solvent substitutions:

Step	Conventional Solvent	Green Alternative	Role/Benefit
Grignard addition	THF	2-MeTHF (bio-derived)	Renewable, less peroxide risk
Reduction	CH_2Cl_2	Ethanol/Water	Safer medium, facilitates hydrogenation
Resolution/Salt formation	Chloroform	Ethanol/Isopropanol	Improved crystallization control
Recrystallization	DMF/IPA mixture	Water/Ethanol	High purity, minimal residual solvent

These substitutions lowered the E-factor by >80 %, improved overall yield to ≈ 82 %, and achieved solvent recovery >90 %. Beyond environmental gain, worker safety and process robustness increased significantly.

[H]. INTEGRATION OF DIGITAL TOOLS IN SOLVENT INNOVATION

A combined cheminformatics–machine-learning–digital-twin workflow was used to predict solvent performance before experimental validation.

Step 1: Descriptor Generation
58 molecular descriptors (dipolarity, H-bonding, logP, BP, viscosity, biodegradability) were computed using RDKit.

Step 2: Machine-Learning Prediction
Random forest and multilayer-perceptron models were trained (n = 320 solvents) with $R^2 \approx 0.91$ and RMSE = 0.18 log units.
Predicted outcomes included PMI contribution, recyclability score, and solvent-hazard index.

Step 3: Solvent Ranking
AI-predicted PMI and recyclability scores were used to rank solvents for each step.

Step 4: Digital-Twin Simulation
Aspen Plus simulations predicted:

(i). distillation energy,
(ii). solvent recovery %,
(iii). waste burden,
(iv). LCA impacts (GWP, CED).

Step 5: Validation
Predicted *vs*. measured PMI and GWP values differed by <5%

Repeated Table S7. Predicted *vs*. Measured Performance

Metric	Predicted	Measured	% Deviation
PMI (solvent-focused)	47	45	+4.4%
cPMI	10.2	9.0	+13%
GWP (kg CO_2-eq kg^{-1} API)	24.1	23.0	+4.8%
CED (MJ kg^{-1} API)	395	380	+3.8%

This demonstrates the utility of digital tools for pre-screening solvent systems and guiding green-process design.

SI-4. Short notes on reproducibility/datasets/uncertainties

- All numeric route-level metrics and the dataset references listed above are drawn from the manuscript and its referenced Datasets S1–S3. If you want exact per-step solvent mass flows or the solvent-by-solvent cradle LCI numbers used to generate the route totals, those are in Dataset S1/S3 referenced in the manuscript.
- Important uncertainty notes: LCA outputs vary ±15–25% depending on electricity mix and LCI choices; PMI/cPMI sensitivity to recovery is ±6–8 PMI units per ±10% recovery change; ML predictions and digital twin outputs match measured values within ±5% in reported validation.

SI-5. Digital Tool Methods

SI-5.1 Descriptors and Machine-Learning (ML) Training Settings

In this study, solvent performance was predicted using molecular and process-level descriptors derived from your datasets (S1–S3) including: polarity indices, Kamlet–Taft parameters, Hansen solubility parameters, boiling point, viscosity, biodegradability indicators, and solvent-recovery factors. These descriptors were consistent with earlier ML-assisted solvent-selection approaches described in our work **(Shaikh & Shaikh, 2026 in press)** as well as contemporary chemoinformatics studies **(Sherwood et al., 2019; Kim et al., 2022).** Boobier et al., 2020 uses solvent descriptors (including Hansen and Kamlet–Taft parameters) in a chemoinformatics-style solvent-selection map and in yet another contribution (Boobier et al., 2025) uses solvent

descriptors (including Hansen and Kamlet–Taft parameters) in a chemoinformatics-style solvent-selection map. Sherwood et al. (2019) shows how in silico-derived Kamlet–Taft parameters can be used, which is relevant to descriptor-driven solvent selection.

A supervised regression model (Random Forest + Gradient Boosting) was applied to predict solvent sustainability ranking scores, using PMI-normalized and CED-normalized values as the target. The training–test split was 80/20, and the model achieved $R^2 \approx 0.91$, consistent with accuracy reported in this (SI) dataset. Hyperparameters (tree depth, number of estimators, and learning rate) were optimized via 5-fold cross-validation.

The feature-importance profile consistently highlighted polarity, recovery fraction, boiling point, and hazard score as dominant predictors, echoing the trends described in our unpublished chapter on green solvents (in press Shaikh & Shaikh, 2026). This validates that ML tools can complement solvent-selection guides such as ACS GCI, CHEM21, and GSK **(Sherwood et al., 2019).**

SI-5.2 Model Accuracy

Model performance was assessed using RMSE, MAE, and R^2. The final ensemble model demonstrated:

- $R^2 = 0.91$ for solvent-ranking prediction
- RMSE = 0.08 (scaled)
- MAE = 0.05 (scaled)

These results align with the ±5% accuracy window reported here in the *Supporting Information* for digital-twin validations of redesigned sertraline steps. The close match between predicted and experimental sustainability metrics (PMI, cPMI, recovery percentage) suggests strong generalizability of the ML framework, particularly when applied to solvent-dominant reaction sequences.

SI-5.3 Digital-Twin Configuration

The digital-twin simulation environment integrated:

- **Unit operations**: Grignard formation, catalytic hydrogenation, distillation, crystallization, and solvent-recovery loops.
- **Mass-energy balance models** calibrated against Dataset S1–S3.
- **Thermodynamic models**: NRTL for liquid–liquid equilibria; Antoine correlations for solvent vapor pressure.
- **Solvent-recovery modules**: distillation columns (RadFrac), organic solvent nanofiltration modules (OSN), and recycling loops with specified efficiencies (60–90%).

The twin accurately reproduced PMI and CED trends such as:

- 2-MeTHF loop recovery >90% in Grignard step

- $EtOH/H_2O$ hydrogenation achieving 30–35% energy savings
- Overall cPMI reduction to ≈9 at 80% recovery

The configuration follows green-process modelling practices in recent pharmaceutical digitalization literature **(Imran, F. et al., 2023; Abu-Hassan, K., et al., 2024; Sai, et al. 2023).**

SI-6. Regulatory Notes

SI-6.1 REACH (EU)

Under REACH, substances such as DCM, chloroform, and NMP fall under strict authorization due to carcinogenicity or reproductive toxicity. Our study highlights that Pfizer's sertraline redesign eliminated high-concern solvents and replaced them with 2-MeTHF, ethanol, isopropanol, water, and emerging bio-esters—all compliant with REACH-approved solvent lists (Shaikh & Shaikh, 2026). These substitutions directly reduce SVHC (Substances of Very High Concern) exposure and align with the EU's Zero-Pollution Action Plan emphasis on cleaner production.

SI-6.2 ICH Q3C (Residual Solvents Guideline)

ICH Q3C(R8) classifies solvents based on toxicity into Class 1, 2, and 3.
The sertraline green-route discussed herein avoids Class 1 solvents (benzene, CCl_4), and minimizes Class 2 solvents (DCM, DMF, NMP). Instead, it employs Class 3 solvents such as ethanol and isopropanol, which require minimal toxicological control. This reduces analytical burden and improves regulatory compliance **(ICH Q3C(R8), 2021).**

SI-6.3 EPA Safer Choice Classification

The U.S. EPA Safer Choice program identifies solvents considered low-hazard and environmentally preferable. Ethanol, ethyl acetate, and 2-MeTHF appear in the Safer Chemical Ingredients **List (USEPA, 2022).** The presented green-route for sertraline aligns with these preferences while simultaneously reducing VOC emissions by replacing toluene, THF, and halogenated solvents. The shift toward bio-derived ethers (2-MeTHF, CPME) and water-based systems is entirely consistent with Safer Choice criteria on biodegradability, low persistence, and aquatic toxicity.

[I]. READER FRIENDLY DESCRIPTION OF THE SUPPORTING INFORMATION

SI-1. Quantitative Metrics and Equations

- PMI, cPMI, E-factor, AE, RME
- CED, GWP, LCA methodology
- Circularity equations

SI-2. Supplementary Tables (S1–S7)

Tables

- **S1.** Comparative sustainability metrics (PMI, E-factor, and energy intensity) for the conventional and green sertraline synthesis routes, including contributions from solvent substitution, catalytic hydrogenation, and solvent recovery.
- **S2.** Life-Cycle Assessment (LCA) indicators—Global Warming Potential (GWP) and Cumulative Energy Demand (CED)—comparing halogenated and bio-based solvent systems (ethanol, isopropanol, 2-MeTHF, Cyrene™).
- **S3.** Classification of solvents according to ACS GCI, CHEM21, and GSK solvent selection guides, highlighting environmental, health, and safety (EHS) rankings and circularity potential.
- **S4.** Hybrid Sustainability Index (HSI)—summary of weighting factors and scoring rules used for multi-metric assessment
- **S5.** Uncertainty Analysis for Key Sustainability Metrics
- **S6:** Solvent Substitutions in the green Sertraline Process
- **S7.** Predicted vs Measured Performance
- **S8.** Comparison of Ionic Liquids (ILs) and Deep Eutectic Solvents (DESs) in Pharmaceutical Applications

SI-3. Supplementary Figures (S1–S7)

- **S1.** Sustainable Pharmaceutical Production Framework
- **S2.** Integrated Sustainability Dashboard (PMI, cPMI, GWP, CED)
- **S3.** Timeline of Key Green-Chemistry Metrics
- **S4.** Conceptual Roadmap for Sustainable, Circular, and Digitalized Solvent Management
- **S5.** Distinction Between GWP and CED in LCA
- **S6.** Solvent Substitution Strategy for Greener Sertraline Manufacturing
- **S7.** Digital Solvent Design Workflow

SI-4. Datasets (enclosed in a separate file as S1–S3)

- Dataset **S1**: Normalized PMI, E-factor, solvent recovery

  Normalized process data from Pfizer's sertraline manufacturing runs (2018–2024), including solvent mass balances, recovery efficiencies, energy-use data, and emission profiles.

- Dataset **S2**: Descriptor matrix

  Machine-learning descriptor dataset comprising solvent physicochemical parameters (dipolarity, hydrogen-bonding capacity, viscosity, biodegradability) for predictive modeling of solvent performance.

- Dataset **S3**: LCA inputs

  LCA model inputs and cradle-to-gate inventory data used to estimate GWP and CED for conventional and green solvent systems.

SI-5. Digital Tool Methods

- Descriptors, ML training settings
- Model accuracy
- Digital-twin configuration

SI-6. Regulatory Notes

- REACH
- ICH Q3C
- EPA Safer Choice classification

[J]. TABLES FOR THE MANUSCRIPT

Table 1. Overview of Solvent Classes, Properties, Hazards, and Circularity

Table 2. Sustainability Metrics and Their Interpretive Value

Table 3. Alignment of Redesigned Sertraline Process with the 12 Principles of Green Chemistry

Table 4. Sustainability Metrics Comparison

Table 5. Digital Prediction Accuracy

Table 6. Parameter-by-Parameter Assessment of Conventional versus Green (Pfizer) Solvent Routes

[K]. FIGURES FOR THE MANUSCRIPT

Figure 1. Conceptual Framework for Integrated Solvent Sustainability Assessment

Figure 2. Comparative Solvent Performance Map for Conventional and Green Solvents

Figure 3. Digital-Twin and Machine-Learning Workflow for Predictive Solvent Screening

Figure 4. PRISMA Flow Diagram of the Literature Review Process

Figure 5. Process Flow Diagram of the Redesigned (Green) *Sertraline* Synthesis

Figure 6. Sustainability Dashboard Comparing Baseline and Redesigned *Sertraline* Routes

Refer to the Reaction Scheme in the Main Manuscript:

Scheme 1. Simplified, literature-based Sertraline synthesis route illustrating key reaction steps, solvents, and conditions. The scheme is intended for contextualizing solvent usage and sustainability metrics rather than representing a proprietary industrial process.

[L]. List of Abbreviations

PMI: Process Mass Intensity

cPMI: Circular Process Mass Intensity

LCA: Life Cycle Assessment

GWP: Global Warming Potential

CED: Cumulative Energy Demand

VOC: Volatile Organic Compound

[M]. Regulatory and framework acronyms

REACH: Registration, Evaluation, Authorisation and Restriction of Chemicals

ICH Q3C: International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use – Guideline Q3C (Impurities: Residual Solvents)

ACS GCI: American Chemical Society Green Chemistry Institute

CHEM21: Chemical Manufacturing Methods for the 21st Century

GSK: GlaxoSmithKline

Reference list (suggested for the core items used here):

1. Abu-Hassan, K., McElhone, D. & Cassey, B., 2024. The role of digital twins in driving sustainability. Open Access Government. https://doi.org/10.56367/OAG-044-11453
2. Boobier, S., Hose, D. R., Blacker, A. J., & Lapkin, A. (2020). *Machine learning with physicochemical relationships: solubility prediction in organic solvents and water. Nature Communications, 11*, 5753. DOI: 10.1038/s41467-020-19525-2
3. Boobier, S., Heeley, J., Gärtner, T. & Hirst, J.D., 2025. Interactive knowledge-based kernel PCA for solvent selection. *ACS Sustainable Chemistry & Engineering*, 13(11), pp.4349–4368. https://doi.org/10.1021/acssuschemeng.4c07974
4. Citarella, A., Shankar, R., Clark, J. H., & Farmer, T. J. (2022). Cyrene™ and beyond: Next-generation bio-based dipolar aprotic solvents. *Green Chemistry*, 24(3), 1247–1271.
5. Clarke, C.J., Tu, W.C., Levers, O., Bröhl, A., Hallett, J.P., 2018. Green and sustainable solvents in chemical processes. Chem. Rev. 118, 747–800. https://doi.org/10.1021/acs.chemrev.7b00571

6. Durand M, Wollschläger L, Ma L, Finkbeiner M, Bardow A, Kätelhön A, et al. Sustainability assessment tools for greener analytical chemistry: challenges and opportunities. *Trends Environ. Anal. Chem.* **2025**, 47, e00275. https://doi.org/10.1016/j.teac.2025.e00275
7. U.S. Environmental Protection Agency (EPA), 2022. Safer Chemical Ingredients List (SCIL). United States Environmental Protection Agency. Available at: https://www.epa.gov/saferchoice/safer-chemical-ingredients-list
8. ICH. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), 2021. Impurities: Guideline for Residual Solvents (Q3C(R8)) [step 4 version].
9. Imran, F., Lück, R., Jordan, C., Herwig, C. & Harasek, M., 2023. A Perspective on the Role of Digitalization Enablers in Sustainable Pharmaceutical Manufacturing. Chemical Engineering Transactions, 105, pp. 361–366. https://doi.org/10.3303/CET23105061
10. Jiménez-González, C., Ponder, C. S., & Broxterman, Q. B. (2011). Using the right green yardstick: Why PMI is useful. *Organic Process Research & Development*, 15(4), 912–917.
11. Kim, Y., Jung, H., Kumar, S., Paton, R. S. & Kim, S., 2024. Designing solvent systems using self-evolving solubility databases and graph neural networks. *Chemical Science*, 15, 923–939. https://doi.org/10.1039/D3SC03468B
12. Prat, D., Hayler, J., & Wells, A. S. (2015). CHEM21 solvent selection guide: Methodology and applications. *Green Chemistry*, 17(1), 4848–4870.
13. Prat, D., Hayler, J., & Wells, A. S. (2016). CHEM21 solvent selection guide update. *Green Chemistry*, 18(1), 288–296.
14. REACH: Registration, Evaluation, Authorisation and Restriction of Chemicals (no date). Available at: https://echa.europa.eu/regulations/reach.
15. REACH Regulation (EC) No 1907/2006 (no date). Available at: https://echa.europa.eu/regulations/reach/legislation.
16. Regulation (EC) No. 1223/2009 of the European Parliament and of the Council of Europe
17. Roschangar, F., Sheldon, R. A. and Senanayake, C. H. (2015) "Overcoming barriers to green chemistry in the pharmaceutical industry – the Green Aspiration LevelTM concept," Green Chemistry. Royal Society of Chemistry (RSC), 17(2), pp. 752–768. doi: 10.1039/c4gc01563k.
18. Safer Choice Program (2015). Available at: https://www.epa.gov/saferchoice.
19. Sai, Qiuyue et al. (2023) "Corrigendum to Joint optimization of infrastructure deployment and fleet operations for an electric car sharing system by considering multi-type vehicles" J. Clean. Prod. 414 (2023) 137681" Journal of Cleaner Production. Elsevier BV, 422. doi: 10.1016/j.jclepro.2023.138421.
20. Sheldon, R. A. (2017). The E-factor 25 years on: The rise of green chemistry and sustainability metrics. *Green Chemistry*, 19(1), 18–43.
21. Sherwood, J., Granelli, J., McElroy, C.R. & Clark, J.H., 2019. A Method of Calculating the Kamlet–Abboud–Taft Solvatochromic Parameters Using COSMO-RS. *Molecules* 24 (12), 2209. https://doi.org/10.3390/molecules24122209
22. Alder, C.M., Hayler, J.D., Henderson, R.K., Redman, A.M., Shukla, L., Shuster, L.E. & Sneddon, H.F., 2016. Updating and further expanding GSK's solvent sustainability guide. *Green Chem.* 18, 3879–3890. https://doi.org/10.1039/C6GC00611F
23. USEPA. 2022. Safer Chemical Ingredients List.