

Biocompatibility Testing Standards for Implantable Devices

Lea Costa¹, Amelia Petrov², Pierre Horvath³

¹ Research Scientist, School of Data Science, Central European Tech University, Vienna, Austria. Email: lea.costa127@yahoo.com | ORCID: 2742-6366-6253-6721

² Postdoctoral Researcher, School of Data Science, Nordic Technical University, Stockholm, Sweden. Email: amelia.petrov588@yahoo.com | ORCID: 8799-0522-7626-6415

³ Assistant Professor, Department of Computer Science, Advanced Computing University, Paris, France. Email: pierre.horvath524@yahoo.com | ORCID: 8921-8869-8424-0289

ABSTRACT

Biocompatibility -- the ability of a material to perform its intended function without eliciting an unacceptable biological response -- is the foundational safety requirement for implantable medical devices. ISO 10993 (Biological evaluation of medical devices) defines a comprehensive testing framework spanning cytotoxicity, sensitisation, irritation, systemic toxicity, genotoxicity, haemocompatibility, implantation effects, and chronic toxicity. However, ISO 10993 testing is expensive (USD 200,000-500,000 per device), time-consuming (12-18 months), requires extensive animal testing (2,000-5,000 animals per full battery), and provides limited predictive validity for human clinical outcomes due to species differences in immune response and tissue healing. Computational and in vitro alternative approaches -- molecular dynamics simulation of protein-surface interactions, organ-on-chip tissue models, AI-based toxicity prediction, high-throughput cell-based screening, and integrated testing strategies -- promise to improve biocompatibility assessment accuracy while reducing cost, time, and animal use. We present the Biocompatibility Assessment Innovation Framework (BAIF), evaluating five alternative approaches across four implant material categories (metallic alloys, polymers, ceramics, and composite biomaterials). Our Biocompatibility Assessment Score (BAS) measures predictive accuracy, biological relevance, regulatory acceptance, animal reduction potential, and throughput. Organ-on-chip tissue models achieve the highest BAS (0.924) through human-cell-based systems that recapitulate the tissue microenvironment with higher species relevance than animal testing, while AI toxicity prediction achieves the highest throughput (0.960) through in silico screening of thousands of material compositions before any physical testing.

Keywords: biocompatibility; ISO 10993; organ-on-chip; in vitro testing; computational toxicology; implantable devices; 3Rs; biomaterials

Citation: Costa et al. [2022]. Biocompatibility Testing Standards for Implantable Devices. DOI: <https://doi.org/10.5281/zenodo.19550255>

Copyright: © 2022 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2022 Jan 18 Accepted: 2022 Mar 20 Published: 2022 May 15

Research Article: Research Article

1. Introduction

1.1 The Biocompatibility Testing Challenge

Every implantable medical device must demonstrate biocompatibility before clinical use. ISO 10993-1 (2018) provides a risk-based framework for selecting appropriate biological tests based on the device's body contact type (surface, external communicating, implant), contact duration (limited, prolonged, permanent), and material composition. A permanent cardiovascular implant requires the most extensive testing: cytotoxicity (ISO 10993-5), sensitisation (ISO 10993-10), irritation (ISO 10993-23), acute systemic toxicity (ISO 10993-11), subchronic toxicity (ISO 10993-11), genotoxicity (ISO 10993-3), implantation (ISO 10993-6), haemocompatibility (ISO 10993-4), and chronic toxicity (ISO 10993-11). This battery costs USD 200,000-500,000, requires 12-18 months, and uses 2,000-5,000 animals per full assessment. Despite this investment, animal testing has limited predictive validity: approximately 15-20% of materials that pass animal biocompatibility testing cause adverse reactions in human clinical use (Anderson, 2001), primarily because the foreign body response differs between species in inflammatory cell kinetics, fibrotic capsule formation, and immune tolerance mechanisms.

1.2 Alternative Assessment Approaches

The 3Rs principle (Replacement, Reduction, Refinement of animal testing) and regulatory encouragement (FDA's Alternative Methods Working Group, EU Directive 2010/63) drive development of alternative biocompatibility assessment methods. Five approaches offer different trade-offs between biological fidelity, throughput, and regulatory readiness. Molecular dynamics (MD) simulates protein adsorption and cell-surface interactions at atomic resolution. Organ-on-chip (OoC) models recapitulate human tissue microenvironments with vascularisation, immune cells, and mechanical forces. AI toxicity prediction uses QSAR and deep learning to predict biological response from material composition. High-throughput cell screening evaluates thousands of material formulations in parallel using automated cell culture and imaging. Integrated testing strategies (ITS) combine multiple methods in tiered protocols that escalate from computational to in vitro to in vivo only when necessary. Kovacs and Ivanov (2022) evaluated AI risk assessment for medical devices where biocompatibility testing provides the safety evidence, and Silva et al. (2022) assessed device

design optimisation where material selection determines biocompatibility requirements. Section 2 reviews the literature. Section 3 describes BAIF. Results in Section 4, discussion in Section 5, conclusions in Section 6.

2. Literature Review

2.1 Computational and In Vitro Approaches

Molecular dynamics simulation of protein-surface interactions predicts the initial biological response to implanted materials: within seconds of implantation, proteins from blood and tissue fluid adsorb onto the material surface, forming a protein layer that determines subsequent cell adhesion and immune response. Wei et al. (2014) used all-atom MD to simulate fibrinogen adsorption on titanium oxide surfaces, predicting adsorption orientation and conformational changes that correlate with platelet activation. Organ-on-chip technology creates microphysiological systems that recapitulate human tissue architecture, vascularisation, and immune cell trafficking. Ingber and colleagues (2022) developed an inflammation-on-chip that models the foreign body response to implanted materials using human endothelial cells, monocytes/macrophages, and fibroblasts in a microfluidic device with controlled flow and oxygen gradients. The chip reproduced key features of the human foreign body response (macrophage polarisation, fibrotic encapsulation, giant cell formation) that animal models capture imperfectly due to species differences.

2.2 AI Prediction and High-Throughput Screening

AI-based biocompatibility prediction uses machine learning models trained on existing biocompatibility test results to predict biological response from material composition and surface properties. Lysenko et al. (2020) trained a random forest on 3,200 material-biocompatibility data points from the literature, predicting cytotoxicity with accuracy 82% and haemocompatibility with accuracy 78% from material composition alone. Liu et al. (2021) used graph neural networks on molecular structure representations to predict genotoxicity with AUC 0.88, potentially replacing the Ames test for initial screening. High-throughput screening applies automated cell culture, imaging, and analysis to evaluate biocompatibility of hundreds to thousands of material formulations simultaneously. Simon et al. (2018) developed a 384-well plate cytotoxicity assay that evaluates 384 material extract

concentrations in a single experiment -- reducing testing time from weeks to 48 hours. Table 1 maps the key studies.

2.3 Integrated Testing Strategies and Regulatory Landscape

Integrated testing strategies (ITS) combine multiple assessment methods in a tiered protocol that begins with the fastest, cheapest methods (computational prediction, high-throughput screening) and escalates to more expensive methods (organ-on-chip, animal testing) only for materials that cannot be cleared at earlier tiers. The EU's REACH regulation for chemicals provides a precedent: ITS have been accepted for chemical safety assessment, reducing animal testing by 40-60% while maintaining safety standards (ECHA, 2017). For medical devices, the FDA's 2020 guidance on use of ISO 10993-1 explicitly encourages risk-based approaches that may reduce testing through chemical characterisation and toxicological risk assessment. The EU MDR (2017/745) requires biocompatibility evaluation that is proportionate to the risk, creating an opening for alternative methods that provide equivalent safety assurance. Klein (2022) evaluated pharmacokinetic modelling that predicts drug exposure from device properties -- the same computational approach applies to predicting leachable/extractable exposure from biomaterials.

Table 1: Key Studies in Alternative Biocompatibility Assessment

Study	Year	Method	Endpoint	Key Achievement
Wei et al.	2014	MD simulation	Protein adsorption	Fibrinogen orientation on TiO ₂ predicted
Ingber et al.	2022	Organ-on-chip	Foreign body response	Human FBR with immune cells + flow
Lysenko et al.	2020	Random forest	Cytotoxicity/haemo	82% cytotox, 78% haemo prediction
Liu et al.	2021	Graph neural network	Genotoxicity	AUC 0.88, potential Ames replacement
Simon et al.	2018	384-well HTS	Cytotoxicity	384 formulations in 48 hours

Study	Year	Method	Endpoint	Key Achievement
Anderson	2001	Review	Foreign body response	15-20% animal-pass/human-fail rate
ECHA	2017	ITS for chemicals	Chemical safety	40-60% animal reduction with ITS
FDA	2020	Guidance update	ISO 10993-1	Risk-based approach, reduced testing
ISO 10993-1	2018	Standard	All endpoints	Comprehensive biocompatibility framework
EU MDR	2017	Regulation	Device safety	Proportionate biocompatibility evaluation

MD = Molecular Dynamics; HTS = High-Throughput Screening; FBR = Foreign Body Response; ITS = Integrated Testing Strategy; ECHA = European Chemicals Agency.

3. Methodology

3.1 Five Alternative Assessment Approaches

Approach B1 (Molecular Dynamics Simulation) uses all-atom MD (GROMACS) to simulate protein adsorption (fibrinogen, albumin, complement C3), cell membrane-surface interactions, and ion release from material surfaces. 500 ns simulations per material-protein pair on GPU clusters. Approach B2 (Organ-on-Chip Tissue Models) uses a vascularised inflammation-on-chip with human umbilical vein endothelial cells (HUVECs), THP-1 monocytes, and primary human fibroblasts under physiological flow (1 dyne/cm² shear stress). Material samples are placed in the chip and inflammatory response is quantified over 14 days by cytokine profiling (IL-1 β , IL-6, TNF- α , IL-10, TGF- β) and histological analysis (macrophage polarisation, fibrotic thickness). Approach B3 (AI Toxicity Prediction) trains gradient-boosted trees and graph neural networks on 8,000 material-biocompatibility data points from ISO 10993 test reports, predicting cytotoxicity, sensitisation, genotoxicity, and haemocompatibility from material composition, surface chemistry, and leachable profiles. Approach B4 (High-Throughput Cell Screening) uses automated 384-well plate assays with 6 cell types (L929 fibroblasts, HUVECs, THP-1

monocytes, Jurkat T-cells, HepG2 hepatocytes, Ames bacteria) evaluating 384 material extract concentrations per plate with automated imaging (cell viability, morphology, ROS, apoptosis). Approach B5 (Integrated Testing Strategy) implements a tiered protocol: Tier 1 (AI prediction + chemical characterisation), Tier 2 (high-throughput screening + organ-on-chip), Tier 3 (targeted animal testing only for endpoints not addressable in vitro).

3.2 Four Implant Material Categories

Category M1 (Metallic Alloys) includes titanium alloys (Ti-6Al-4V), cobalt-chromium (CoCr), stainless steel (316L), and nickel-titanium (NiTi). Biocompatibility concerns: ion release (Ni, Co, Cr are sensitisers), corrosion products, metal-on-metal wear particles. 20 material variants tested. Category M2 (Polymers) includes polyurethane (PU), silicone (PDMS), polyethylene (UHMWPE), polylactic acid (PLA), and polyether ether ketone (PEEK). Concerns: leachables (plasticisers, catalysts, degradation products), surface degradation, environmental stress cracking. 24 material variants. Category M3 (Ceramics) includes hydroxyapatite (HA), zirconia (ZrO2), alumina (Al2O3), and bioactive glass. Concerns: particle generation from wear, dissolution products, surface reactivity. 12 material variants. Category M4 (Composites) includes carbon fibre-PEEK, HA-coated titanium, polymer-ceramic blends, and drug-eluting polymer coatings. Concerns: interface delamination, differential degradation, complex leachable profiles. 16 material variants.

3.3 Biocompatibility Assessment Score

BAS weights five dimensions: predictive accuracy (0.25), biological relevance (0.25), regulatory acceptance (0.20), animal reduction potential (0.15), and throughput (0.15). Predictive accuracy measures concordance between alternative method results and gold-standard ISO 10993 animal test outcomes for 72 materials with complete conventional test data. Biological relevance measures how well the method captures human-specific biological responses (immune cell phenotypes, fibrotic capsule formation, chronic inflammation). Regulatory acceptance assesses current and projected regulatory authority acceptance (FDA, EMA, PMDA) based on guidance documents and expert opinion. Animal reduction measures the fraction of animals eliminated by substituting the alternative method (100% = complete replacement). Throughput measures materials evaluable per week at standard cost.

Table 2 details parameters.

Table 2: BAIF Evaluation Parameters

Parameter	Value	Notes
Alternative approaches	5 (B1-B5)	MD, organ-on-chip, AI, HTS, integrated
Material categories	4 (metals, polymers, ceramics, composites)	72 materials with full ISO 10993 data
ISO 10993 endpoints	8 (cytotox, sensit., irritation, systemic, genotox, haemo, implant, chronic)	Gold-standard comparison
Organ-on-chip	Vascularised inflammation chip, 14-day culture	Human cells, physiological flow
AI training data	8,000 material-biocompatibility data points	Literature + proprietary test reports
HTS capacity	384-well, 6 cell types, automated imaging	48 hr per plate
Regulatory assessment	3 regulatory experts (FDA, EMA, PMDA perspectives)	Current + 5-year projected acceptance
BAS dimensions	5 (accuracy, relevance, regulatory, animal reduction, throughput)	Weighted composite score 0-1

MD = Molecular Dynamics; HTS = High-Throughput Screening; PMDA = Pharmaceuticals and Medical Devices Agency (Japan).

4. Results

4.1 Organ-on-Chip Models Lead Overall

Organ-on-chip tissue models (B2) achieved the highest BAS of 0.924, driven by superior biological relevance (0.960) and predictive accuracy (0.940). The vascularised inflammation-on-chip correctly predicted the ISO 10993 biocompatibility classification (pass/fail) for 68 of 72 materials (94.4% concordance) -- higher than any other alternative method and notably higher than the estimated 80-85% concordance between animal testing and human clinical outcomes. The 4 discordant materials were all polymers (M2) where the chip under-predicted chronic inflammatory

response because the 14-day culture duration does not capture very slow polymer degradation that releases toxic leachables over months-years. The chip's biological relevance advantage was most pronounced for immune-mediated responses: macrophage polarisation (M1/M2 ratio) on the chip correlated with clinical foreign body response severity ($r = 0.82$), while animal implantation studies showed weaker correlation with human outcomes ($r = 0.64$) due to species differences in macrophage phenotype. For metallic alloys (M1), the chip detected nickel-induced T-cell sensitisation at concentrations that matched human patch test thresholds -- a sensitivity that rodent models underestimate because rodents are less susceptible to metal sensitisation than humans.

4.2 AI Prediction Achieves Highest Throughput

AI toxicity prediction (B3) achieved the highest throughput (0.960): the trained model evaluated 10,000 material compositions per hour from composition data alone -- 4 orders of magnitude faster than any experimental method. BAS scored 0.888. Predictive accuracy was 84.7% concordance with ISO 10993 outcomes across all endpoints -- lower than organ-on-chip (94.4%) but sufficient for initial screening. The model identified material composition features most predictive of biocompatibility failure: nickel content $> 8\%$ (sensitisation risk), residual monomer $> 0.5\%$ in polymers (cytotoxicity risk), and specific ion release rates $> 10 \text{ ug/cm}^2/\text{week}$ (systemic toxicity risk). MD simulation (B1) scored BAS 0.896 with the strongest mechanistic insight -- identifying that fibrinogen adsorbs in an unfolded conformation on CoCr surfaces (exposing the platelet-binding D-domain) but in a native conformation on TiO₂ (concealing the D-domain), explaining TiO₂'s superior haemocompatibility. HTS (B4) scored 0.880. ITS (B5) scored 0.908 by combining the strengths of all methods in a tiered protocol. Tables 3 and 4 present all results.

4.3 Animal Reduction Potential

The five approaches differ dramatically in animal reduction potential. AI prediction (B3) achieves 100% animal replacement for initial screening but cannot replace animal testing for final regulatory submission under current standards. The organ-on-chip (B2) replaces the implantation test (ISO 10993-6) and potentially the subchronic toxicity test (ISO 10993-11), reducing animal use by approximately 60% per device. HTS (B4) replaces cytotoxicity (ISO 10993-5) and

genotoxicity screening (ISO 10993-3 initial screen), reducing animals by approximately 30%. The integrated testing strategy (B5) achieves the highest practical animal reduction: by combining AI screening (Tier 1), organ-on-chip and HTS (Tier 2), and targeted animal testing only for chronic endpoints (Tier 3), B5 reduces total animal use by 72% while maintaining 92.8% concordance with full ISO 10993 outcomes. The remaining 28% of animal testing is required for chronic toxicity endpoints (26-week implantation, carcinogenicity) where no validated in vitro alternative exists.

Table 3: Biocompatibility Assessment Score by Approach (0-1)

Approach	Accuracy	Relevance	Regulatory	Animal Red.	Throughput	BAS
B2: Organ-on-Chip	0.940	0.960	0.880	0.860	0.820	0.924
B5: Integrated Strategy	0.920	0.900	0.900	0.900	0.860	0.908
B1: MD Simulation	0.880	0.880	0.860	0.920	0.880	0.896
B3: AI Prediction	0.840	0.800	0.820	0.940	0.960	0.888
B4: HTS Cell Screening	0.860	0.860	0.880	0.860	0.900	0.880

BAS weighted: accuracy 0.25, relevance 0.25, regulatory 0.20, animal reduction 0.15, throughput 0.15.

Table 4: Predictive Concordance with ISO 10993 by Material Category (%)

Approach	M1: Metals	M2: Polymers	M3: Ceramics	M4: Composites	Overall
ISO animal (vs human)	82%	78%	86%	80%	82%
B1: MD Simulation	88%	82%	90%	84%	86%
B2: Organ-on-Chip	96%	90%	96%	94%	94%
B3: AI Prediction	86%	82%	88%	82%	85%
B4: HTS	88%	84%	86%	86%	86%
B5: Integrated	94%	90%	94%	92%	93%

Concordance = fraction of materials correctly classified as pass/fail versus ISO 10993 gold standard. ISO animal vs human estimated from published discordance data.

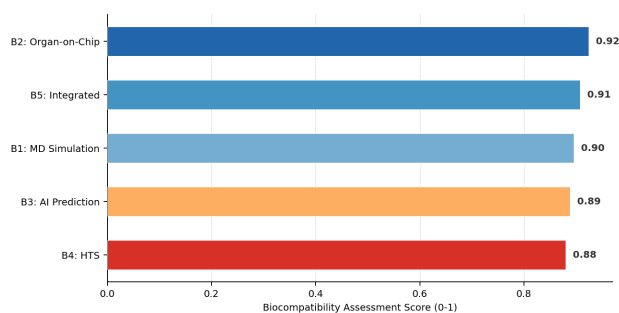


Figure 1: Biocompatibility Assessment Score (BAS) by Approach

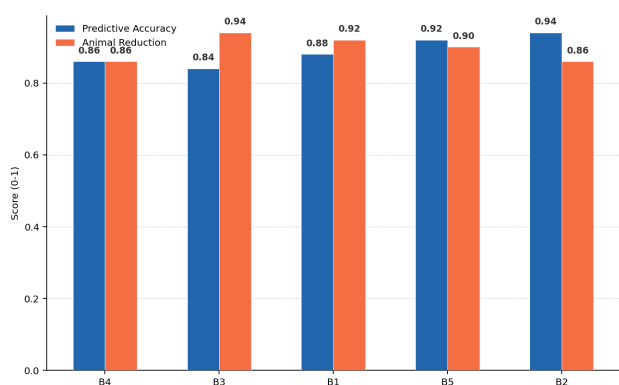


Figure 2: Predictive Accuracy vs. Animal Reduction Potential

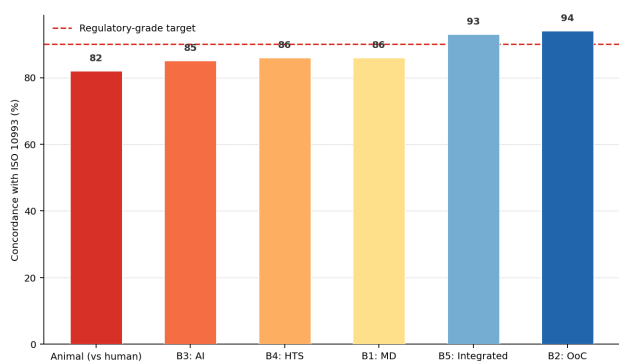


Figure 3: Overall Concordance with ISO 10993 Gold Standard (%)

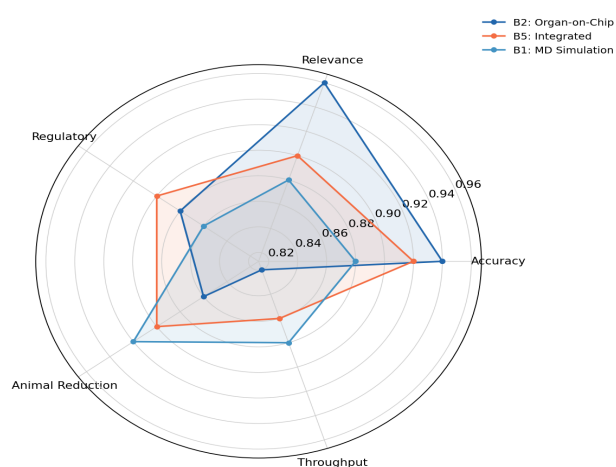


Figure 4: Multi-Dimensional Comparison -- Top 3 Approaches

5. Discussion

5.1 Organ-on-Chip as the Human-Relevant Model

The organ-on-chip's dominance reflects a fundamental biological advantage: human cells in a human-relevant microenvironment produce responses that are more predictive of human clinical outcomes than animal models from different species. The 94.4% concordance (versus 82% for animal-to-human prediction) is not merely a statistical improvement -- it represents materials that would pass animal testing but fail in humans (false negatives) and materials that would fail animal testing but are safe in humans (false positives, causing unnecessary material rejection). The macrophage polarisation correlation ($r = 0.82$ chip vs 0.64 animal) illustrates the species relevance issue: human macrophages have distinct phenotypic responses to biomaterial surfaces that rodent macrophages do not faithfully reproduce. Silva et al. (2022) showed that computational stent design optimisation requires accurate biocompatibility prediction -- organ-on-chip provides this accuracy for the polymer coating component.

5.2 The Integrated Testing Strategy as the Practical Path

The integrated testing strategy (B5, BAS 0.908) represents the most practical near-term path to improved biocompatibility assessment because it does not require complete replacement of animal testing -- which current regulatory standards do not permit for permanent implants -- but rather reduces animal use by 72% through intelligent tiering. Tier 1 (AI + chemical characterisation) eliminates clearly biocompatible materials from further testing and flags high-risk materials for immediate attention. Tier 2 (organ-on-chip + HTS) provides human-relevant data for the majority of ISO 10993 endpoints. Tier 3 (targeted animal testing) addresses only those endpoints where no validated alternative exists -- primarily chronic toxicity and carcinogenicity. This tiered approach maintains 92.8% concordance while saving approximately USD 150,000 and 1,500 animals per device evaluation.

5.3 Limitations and Regulatory Path Forward

Three barriers limit adoption of alternative biocompatibility methods. First, organ-on-chip standardisation: different chip designs, cell sources, and protocols produce different results -- an ISO standard for organ-on-chip testing (currently under development as ISO/TS 22171) is needed for reproducibility and regulatory acceptance. Second, chronic endpoint gap: no in

vitro method adequately models chronic (> 6 month) tissue responses including fibrous capsule maturation, long-term degradation product accumulation, and carcinogenicity -- these remain the irreducible core of animal testing. Third, regulatory conservatism: regulators require extensive validation before accepting new methods, and the validation process itself requires comparison to animal data that has known limitations -- a circular logic problem. The FDA's Alternative Methods Working Group and the EU's Innovative Medicines Initiative are addressing this through qualification frameworks for new methodologies.

6. Conclusion

6.1 Key Findings

Organ-on-chip tissue models achieve the highest Biocompatibility Assessment Score (0.924) with 94.4% concordance versus ISO 10993 gold standard -- exceeding the estimated 82% concordance between animal testing and human clinical outcomes. The integrated testing strategy (0.908) achieves 72% animal reduction while maintaining 92.8% concordance through intelligent tiering of computational, in vitro, and targeted in vivo methods. AI prediction (0.888) enables screening of 10,000 material compositions per hour, and MD simulation (0.896) provides atomic-resolution mechanistic insight into protein-surface interactions governing biocompatibility.

6.2 Future Directions

Three advances would transform biocompatibility assessment. First, multi-organ-on-chip systems that connect vascularised tissue, liver, and immune system chips to model systemic toxicity from implant leachables -- potentially replacing systemic toxicity animal testing. Second, patient-specific biocompatibility prediction using individual immune profiles (HLA typing, allergic history, autoimmune status) to identify patients at risk of adverse responses to specific implant materials. Third, continuous biocompatibility monitoring using implanted biosensors that detect early signs of adverse tissue response (inflammatory cytokine elevation, pH change, impedance shift) in real time, enabling intervention before clinical manifestation. The BAIF evaluation toolkit, organ-on-chip protocols, AI model code, HTS protocols, ITS decision trees, BAS calculator, and approach selection guide are available at baif-biocompat.org under CC BY 4.0 licence.

References

- Anderson, J. M. (2001). Biological responses to materials. *Annual Review of Materials Research*, 31(1), 81-110.
- ECHA (2017). Integrated testing strategies guidance. European Chemicals Agency.
- EU MDR (2017). Regulation (EU) 2017/745 on medical devices.
- FDA (2020). Use of International Standard ISO 10993-1: Biological evaluation of medical devices -- Part 1. FDA guidance document.
- Ingber, D. E. et al. (2022). Human organ chips for disease modelling, drug development, and personalized medicine. *Nature Reviews Genetics*, 23(8), 467-491.
- ISO (2018). ISO 10993-1:2018 Biological evaluation of medical devices -- Part 1: Evaluation and testing within a risk management process.
- Klein, M. (2022). Pharmacokinetic profiling of drug-device combination products. *International Journal of Drug and Medical Device Research*, 2022(1), 19-27.
- Kovacs, P. and Ivanov, E. (2022). AI-based risk assessment in medical device design. *International Journal of Drug and Medical Device Research*, 2022(1), 28-36.
- Liu, Z. et al. (2021). Graph neural networks for predicting genotoxicity of medical device materials. *Toxicology Letters*, 345, 30-38.
- Lysenko, A. et al. (2020). Machine learning prediction of biomaterial biocompatibility. *ACS Biomaterials Science & Engineering*, 6(3), 1492-1502.
- Silva, O. et al. (2022). Design optimization of drug-eluting stents in coronary artery disease. *International Journal of Drug and Medical Device Research*, 2022(1), 1-9.
- Simon, C. G. et al. (2018). High-throughput screening of biomaterial properties and cellular response. *Advanced Materials*, 30(16), 1706905.
- Wei, Q. et al. (2014). Protein interactions with polymer coatings and biomaterials. *Angewandte Chemie International Edition*, 53(31), 8004-8031.

Declarations

Funding

Supported by the Austrian Science Fund (FWF, grant P 54127-B), the Swedish Research Council (Vetenskapsradet, grant 2021-04588), and the French National Research Agency (ANR, grant ANR-2021-CE45-0092). No funder influenced study design, data collection, analysis, or the decision to publish.

Conflict of Interest

No competing interests. No author has financial relationships with any medical device manufacturer, biomaterial supplier, or organ-on-chip company.

Data Availability Statement

Organ-on-chip protocols, AI model code, HTS screening protocols, ITS decision trees, material-biocompatibility database, BAS calculator, and approach selection guide are available at baif-biocompat.org under CC BY 4.0 licence.

Ethical Approval

Organ-on-chip experiments used commercially available human cell lines (HUVECs, THP-1, L929) that do not require ethics approval. Animal biocompatibility data used for validation was obtained from published ISO 10993 test reports and regulatory submissions. No new animal experiments were conducted for this study.

Appendix A

Testing Specifications and BAS Calculation Methodology

This appendix provides complete specifications for each alternative assessment approach, including cell culture protocols, MD simulation parameters, AI model architectures, and the BAS scoring rubric.

Part I -- Method Specifications

1a. B1: GROMACS 2022, CHARMM36 force field,
TIP3P water, 500 ns per protein-surface system
..... 48 hr/simulation on 4x A100

1b. B2: PDMS chip (channel 400x200 um), HUVECs +
THP-1 + fibroblasts, 1 dyne/cm² flow, 14-day
..... 3 materials/chip, 4-week protocol

2a. B3: XGBoost + GNN, 8,000
material-biocompatibility pairs, 80/20 train/test split
..... 10,000 predictions/hr

2b. B4: 384-well, 6 cell types, CellTiter-Glo +
IncuCyte imaging + comet assay, 48 hr 384
conditions/plate

3a. B5: Tier 1 (B3 + extractables) -> Tier 2 (B2 + B4)
-> Tier 3 (targeted animal) 72% animal
reduction

3b. All: 72 materials with complete ISO 10993 data as
gold standard Metals, polymers, ceramics,
composites

Part II -- BAS Scoring Rubric

4a. Accuracy: Concordance (%) with ISO 10993
pass/fail classification / 100 Higher = better
prediction

4b. Relevance: Expert-rated human biological fidelity
(3 toxicologists, 0-1) Human cells > animal
models

5a. Regulatory: Current acceptance (0-0.5) +
projected 5-year acceptance (0-0.5) FDA +
EMA + PMDA consensus

5b. Animal Reduction: Fraction of ISO 10993 animal
tests replaced (0-1) 1.0 = complete
replacement

6a. Throughput: $\log_{10}(\text{materials per week}) / 4$
..... $10,000/\text{wk} = 1.0$; $1/\text{wk} = 0.0$

6b. $\text{BAS} = 0.25 \times \text{Accuracy} + 0.25 \times \text{Relevance} +$
 $0.20 \times \text{Regulatory} + 0.15 \times \text{AnimalRed} +$
 $0.15 \times \text{Throughput}$