Mathematical modeling reveals differential effects of erythropoietin on proliferation and lineage commitment of human hematopoietic progenitors in early erythroid culture

Daniel Ward,[1] Deborah Carter,[2] Martin Homer,[1] Lucia Marucci,[1] and Alexandra Gampel[2]

[1]Department of Engineering Mathematics, Faculty of Engineering, University of Bristol; and [2]School of Biochemistry, Faculty of Medical and Veterinary Science, University of Bristol, UK

 doi:10.3324/haematol.2015.133637

Received: July 13, 2015.

Accepted: November 18, 2015.

Pre-published: November 20, 2015

Correspondence: a.gampel@bristol.ac.uk

Mathematical modelling reveals differential effects of erythropoietin on proliferation and lineage commitment of human hematopoietic progenitors in early erythroid culture.

Supplementary Methods

Erythroid culture

PBMCs were washed 4-6 times in Hanks Balanced Salt Solution + citrate dextrose solution 0.6% (v/v) at low centrifugation speed (200g, 10 min) to remove platelets. Red cells were removed by lysis in 155 mM NH4Cl + 0.1 mM EDTA + 1 mM KHCO3. CD34+ cells were isolated using the Ultrapure Human CD34+ isolation kit (Miltenyi). Isolated cells were cultured in erythroid medium (EM); Stem Span (Stem Cell Technologies) + 10 ng/ml Stem Cell Factor (SCF) (for Day 0 to Day 4) and 50 ng/ml (for Day 5 to Day 11), 40 ng/ml Insulin-like growth factor1 (IGF-1),1 ng/ml interleukin-3 (IL-3), 1 mM dexamethasone and with or without 2 U/ml erythropoietin (epo). SCF, IL-3 and IGF-1 were recombinant human proteins purchased from Miltenyi Biotec. For CD36 depletion studies, PBMCs were enriched for $CD34^+$ populations by lineage depletion (Miltenyi Biotec human lineage depletion kit). For colony assays, 450 cells were plated in duplicate in StemMACS HSC-CFU medium complete with epo (Miltenyi).

Flow cytometry and fluorescence assisted cell sorting (FACS)

Cells were stained with anti-human-specific antibodies at a dilution of 1:10 in PBS + 1% BSA for 20 minutes, then washed in PBS and analysed on the MACSQuant. Compensation was done using ecomp beads (ebioscience) and the automated compensation multi-colour tool on the MACSQuant. Post-acquisition analysis was done with FlowJo v7.6.5 for proliferation analysis and FlowJo vX0.7 for all other analysis. Live cells were gated based on scatter properties and on the absence of Propidium iodide uptake, except for CFSE-stained cells where live cells were gated solely on scatter properties. Fluorescence was visualised with BiExponential scaling and positive gates were set using isotype controls. For Fluorescence Activated Cell Sorting (FACS), 2-5 × 10^5 cells cultured for 2-4 days were stained with antibodies to CD34 and CD36. Unstained control and single stained beads were used to set scatter gates and for compensation. $CD34^+CD36^-$ and $CD34^+CD36^+$ cells were sorted using a BD Influx Cell Sorter and sort purity was determined by analysis on the MACSQuant.

$CD36^+$ cell depletion

A $CD34^+$-enriched population was generated by lineage depletion (Miltenyi Biotec) of PBMNC. Cells were cultured for 2, 4 or 6 days, labelled with biotin-conjugated CD36 antibody and $CD36^+$ cells were depleted using the Miltenyi MACS microbead technology with streptavidin microbeads. Cells were analysed by flow cytometry before and after depletion.

Identification of the mathematical model

To fit unknown parameters of the mathematical model to data, we minimised the root mean square error (RMSE), defined as sample standard deviation of the differences between simulated values, and training data. The training data (dataset A, Table S1) consists of CFSE tracking of CMP and MEP generational cell counts (Figures 5A, S3A and S3B). We initially ran a hybrid simulated annealing algorithm (simulannealbnd solver in the MATLAB optimtool routine, combined with pattern search performed at the end of the simulated annealing) upwards of 30 times in a range of initial values and bounds (the bounds were set at [0 1] for all rates, initial estimation ± 2 hours for delays and initial estimation ± 4 hours for τ_C and τ_M, with initial parameter estimations suggested by the experimental data where possible). Best fittings (in terms of lowest RMSE) were selected to further refine the

parameter estimations, and optimise the bounds, via the MATLAB fmincon routine. The options specified for the parameter identification were left at the default values in both optimtool and fmincon. We explored the effect of error in both target fitting data and initial parameter estimation is some detail; see below.

Model reduction

The system was fitted to experimental data (dataset A, Table S1) initially in the most general case of full heterogeneity; i.e. all 37 parameters (Table 1) are generation-dependent. This provided insight as to the general system behaviour and allowed us to simplify the model. Firstly, it appeared that the system is reasonably homogenous for MEP proliferation and heterogeneous for CMP proliferation, but homogenous for the delays between the '*b*' and '*a*' compartments for all generations of CMP, and G(4) onward of MEP. This suggests the use of a single delay for CMP and three delays for MEP (defining $\Gamma_{i\geq 4} = \Gamma_3$). Also, it was clear that the influence of maturation on the system dynamics was primarily from G(0), since maturation rates for later generations, μ_{1-5}, are small (.004, Table S2) compared with the maturation rate of G(0), μ_0, (.016, Table S2).To account for the small effect of maturation in later generations, we defined $\mu_{i\geq 1} = \mu_{\mathrm{F}}$ as the final maturation rate of G(0). The parameters that govern the dynamics of the generalised logistic functions ($Q_\alpha, R_\alpha, Q_\delta, R_\delta, Q_\mu, R_\mu$) were determined using dataset A (CFSE tracking of CMP and MEP generational cell counts in Figures 5A, S3A and S3B) and then remain fixed throughout. These simplifications reduced the system to 20 parameters: 12 proliferation rates (α_{0-5}, δ_{0-5}), 2 maturation rates (μ_{I}, μ_{F}), 4 compartmental delays (Δ, Γ_{1-3}) and 2 proliferation initiation delays (τ_{C}, τ_{M}). The reduction in the number of free parameters allows the optimisation algorithms to converge much more quickly, while still providing a very good fit to experimental data.

Model validation

The model was validated using a second, independent experimental data set ('dataset B', Table S1) in two different ways. First we used the parameter values obtained for dataset A, and used the model to predict time-courses for dataset B, starting from the initial values of dataset A (Figures S5A and S5B, red lines). The total MEP cell count shows very good quantitative agreement, while there is some deviation in the total CMP cell count (Figure S5C). However, it is important to note that the absolute number of CMP cells is significantly smaller than for MEP (Figure S5C). The RMSE calculated from dataset B was 355.02 (4.07%) for CMP and 412.87 (4.26%) for MEP, which although larger than the fitted data set, is still good. Secondly, we re-ran the optimisation routine to find the parameters for dataset B: parameters are reported in Table 2, and fitting results are plotted in Figures S5A-S5C (black lines). Re-optimisation lowered the RMSE to 150.29 (2.09%) for CMP and 218.6 (3.32%) for MEP, and resulted in only small changes to parameter values, indicating high confidence in the model and parameter values.

Robustness analysis

To test the robustness of the parameter identification methods used, and to quantify the sensitivity of the model to measurement error and also initial parameter estimation error, further *in-silico* experiments were conducted on sets of 1000 'simulated data' points. This simulated data was generated by adding independent normally distributed random errors to known solutions of the DDE system, allowing us test the system over a range of increasing errors. The results of these experiments are shown in Table S3. The initial tests consisted of adding error to the known solution such as to make it comparable with the experimental data, using this 'simulated data' to identify how far from the known solution the parameters shift. The addition of 3.07% measurement error (order of magnitude of the measured experimental data) results in a mean shift of 0.829% in

parameters (Table S3, analysis (a)),, and an indistinguishable qualitative change in timecourses. Increasing the simulated measurement error to 10% results in a shift of only 2.14% from the experimentally-derived parameters (Table S3, analysis (c)). Furthermore, we also investigated the effect of combining simulated measurement error with initial parameter estimation error (presenting a worst case scenario, as in practice parameter estimation is aided by visual investigation of qualitative properties, not applied in these tests). Errors of 10% measurement and 10% estimation error result in a shift in parameters of only 2.95% (Table S3, analysis (g)). The extensive datasets used for these *in silico* experiments provided us with a clearer understanding of the local parameter space of our DDE system. In this worst case scenario the resulting parameter identification still gives quantitatively and qualitatively accurate fits to the experimental data set (Figures 5A, S3A and S3B). To garner a more extensive view of the sensitivity of the parameter optimisation to errors in the initial parameter estimate, we tested a further set of 600 *in silico* experiments, increasing the initial parameter estimation error for experimentally comparable simulated data (Table S3, analysis h-k). In this case increasing the initial parameter estimation error from 4% to 16% results in a quantitative change in data fit from 99 cells to 168 cells; an increase in 0.68% in fitting error. We thus conclude that our parameter optimisation methods are robust to noise, both in experimental measurement and initial parameter estimates.

Extrapolation

To predict cell counts beyond T90, where we do not have CMP and MEP generational cell count data (datasets E & F, table S1), we extrapolate by fitting a linear relationship to the proliferation rates obtained above. We also tested alternative schemes, including assuming proliferation rates are constant (with a range of different values tested), that they saturate (again, various levels were tested), and that rates may peak and decrease with time. In each of these cases the longer time-course simulations do not yield vastly better fits, considering they involve assuming more complicated parameter dynamics which there is insufficient data to justify. Cases where the parameters are increasing with each generation do on the whole fit the later stage cell counts more closely. It should be noted that each alternative scheme does yield a fit that falls within the experimental error and as such it is possible that the later stage dynamics (T90-T186) are less sensitive to proliferation rate than the earlier stages (up to T90).

Table S1. Experimental data summary. Summary details of all data sets: descriptive label, conditions (±epo), type of data collected, total time course of experiment, data point times, number of independent experiments, and figures in which results are reported. Datasets E and F are the averages of experiments E-1, E-2 and F-1,F-2, respectively.

Dataset label	Conditions	Data collected	Timecourse	Timepoints (total, initial, interval)	n	Figures
A	+epo	CMP, MEP generational counts. Total cell counts.	90 hours	total $= 11$, $t_0 = 10$ h, $\Delta t = 8$ h	3	5, S4
B	+epo	CMP, MEP generational counts. Total cell counts.	90 hours	total $= 11$, $t_0 = 10$ h, $\Delta t = 8$ h	3	S5
C	+epo	CMP, MEP generational counts. Total cell counts.	90 hours	total $= 4$, $t_0 = 18$ h, $\Delta t = 24$ h	3	6C, 6F, S6 A and B
D	−epo	CMP, MEP generational counts. Total cell counts.	90 hours	total $= 4$, $t_0 = 18$ h, $\Delta t = 24$ h	3	6C, 6F, S6 A and B
E	+epo	Total cell counts.	186 hours	total $= 8$, $t_0 = 18$ h, $\Delta t = 24$ h	6 (E-1/, E-2)	6C, 6E, S6C
F	−epo	Total cell counts.	186 hours	total $= 8$, $t_0 = 18$ h, $\Delta t = 24$ h	6 (F-1/, F-2)	6C, 6E, S6C

Table S2. Parameter optimisation for the full (37 parameter) model. Parameters fitted to the full original 37 parameter model, forming the basis for the assumptions made about system homogeneity and heterogeneity that allowed for simplifications that reduced the system to the 20 parameter case. Mean change in parameter fitting is 0.281%

Parameter	20 parameter	37 parameter	% Difference
α_0	0.018	0.0183	-1.552
α_1	0.054	0.0541	-0.2030
α_2	0.108	0.1077	0.2322
α_3	0.181	0.1805	0.2896
α_4	0.213	0.213	0.0487
α_5	0.195	0.195	-0.0710
μ_i	0.016	0.0165	-3.2250
μ_f	0.004	0.00393	1.7078
μ_1	0.004	0.004004	-0.1030

μ_2	0.004	0.004004	-0.1030
μ_3	0.004	0.004004	-0.1030
μ_4	0.004	0.004004	-0.1030
μ_5	0.004	0.004003	-0.0741
δ_0	0.128	0.127	0.4944
δ_1	0.123	0.123	-0.2008
δ_2	0.119	0.119	0.1280
δ_3	0.127	0.127	0.2680
δ_4	0.129	0.129	0.1776
δ_5	0.11	0.1098	0.1675
τ_c	30.394	30.397	-0.0106
τ_m	27.705	27.702	0.0102
Δ_1	9.8	9.790	0.0978
Δ_2	9.8	9.803	-0.0323
Δ_3	9.8	9.805	-0.0561
Δ_4	9.8	9.762	0.3858
Δ_5	9.8	9.805	-0.0547
Γ_1	14.549	14.55	-0.0270
Γ_2	9.817	9.823	-0.0692
Γ_3	8.017	8.024	-0.0981
Γ_4	8.017	8.020	-0.0437
Γ_5	8.017	8.001	0.1929
Q_α	0.58	0.5801	-0.0253
R_α	0.44	0.4400	-0.0033
Q_μ	0.618	0.6179	0.0069
R_μ	0.065291	0.0653	-0.0178
Q_δ	0.56	0.5599	0.0057
R_δ	0.43	0.4299	0.0129

Table S3. Results from optimisation robustness and system sensitivity analysis. To test the optimisation robustness of the system 7 cases were tested (cases (a)-(g)) with increasing simulated data error (DE) which is comparable to the experimental error and increasing initial parameter guess error (GE) to analyse how a less accurate prediction can still reach the same solution. To test system sensitivity 4 cases (cases (h)-(k)) were tested increasing the GE and measuring the data fitting error.

Analysis	Parameter Shift (%)	Fitting Error (#cells)
(a) ±3.07% DE	0.82	75.83
(b) ±5% DE	0.92	102.04
(c) ±10% DE	2.14	205.03
(d) ±3.07% DE & ±5% GE	0.79	102.88
(e) ±3.07% DE & ±10% GE	1.12	114.43
(f) ±10% DE & ±5% GE	3.17	222.14
(g) ±10% DE & ±10% GE	2.95	230.95
(h) ±4% DE & ±4% GE	3.13	99.21
(i) ±4% DE & ±8% GE	2.74	105.2
(j) ±4% DE & ±12% GE	3.66	149.7
(k) ±4% DE & ±16% GE	3.25	167.8

Gampel, Figure S1

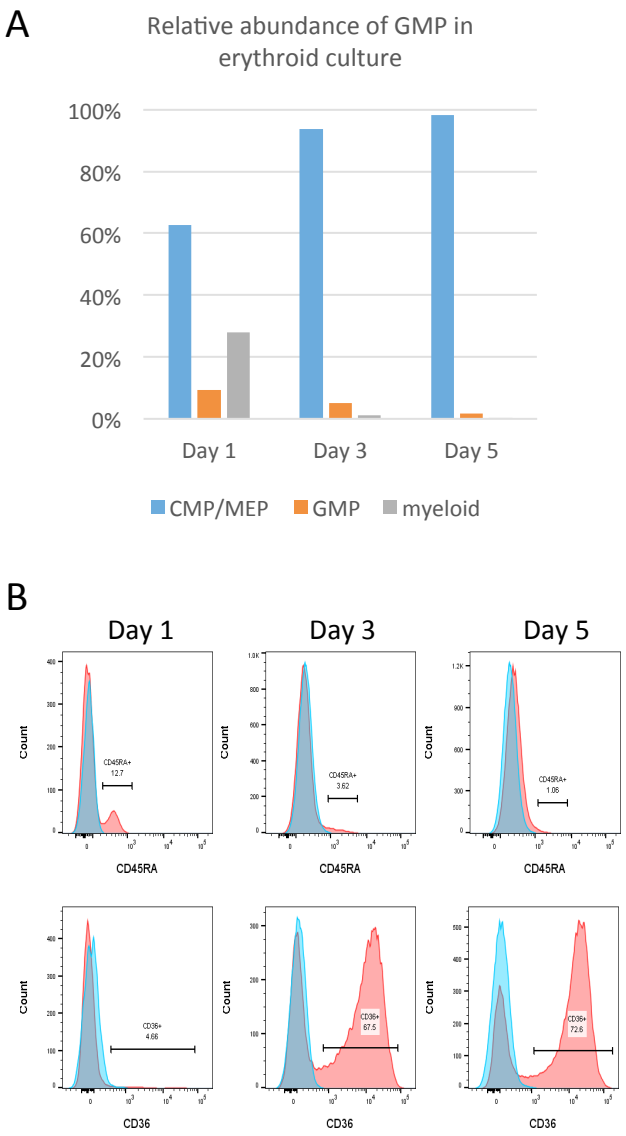

Figure S1. CMP do not give rise to significant numbers of GMP in erythroid culture. CD34$^+$ cells were cultured under standard conditions in the absence of dexamethasone for 5 days and analysed on Days 1, 3 and 5. **(A)** The relative abundance of CMP (CD34$^+$ CD45RA$^-$), GMP (CD34$^+$ CD45RA$^+$) and myeloid cells (CD34$^-$CD45RA$^+$) is shown. **(B)** Populations gated on CD34$^+$ were analysed for CD45RA (top panels) and CD36 (lower panels) surface levels. The population of CD45RA$^+$ cells at Day 1 (13%) is reduced to 1% by Day 5. By contrast, the CD36$^+$ population increases dramatically from Day 1 to Day 5.

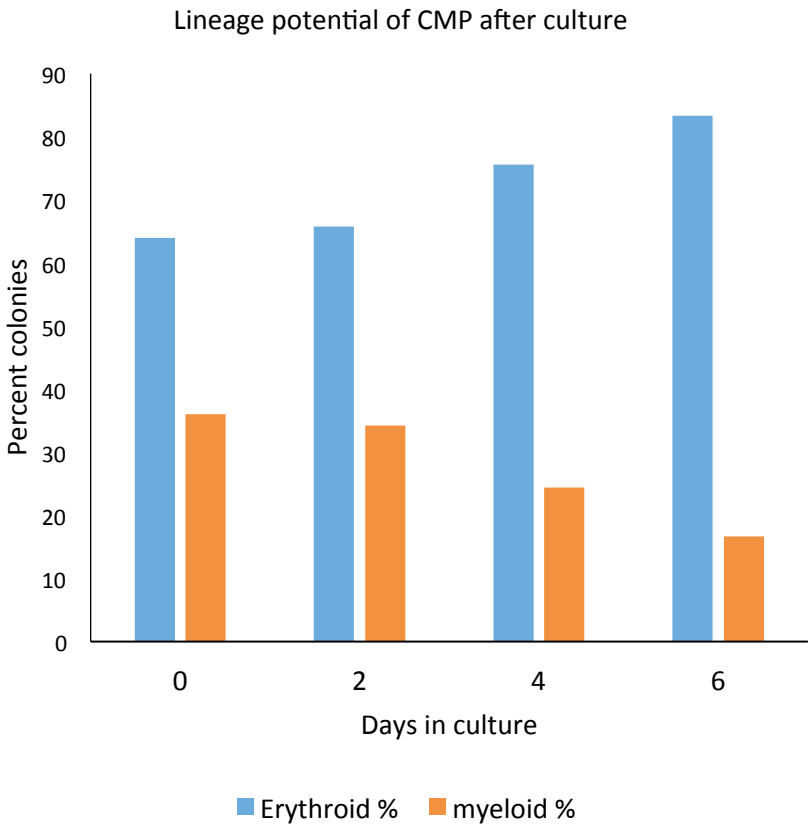

Figure S2. **CD36$^-$ cells retain multipotency in culture**. CD36$^+$ cells were removed from erythroid cultures by immuno-magnetic separation every 2 days from Day 2 and CD36$^-$ cells were plated in methylcellulose immediately after separation. CD36$^-$ cells continue to produce both erythroid (blue) and myeloid (orange) colonies even after 6 days in culture although the erythroid:myeloid colony ratio decreases.

Gampel, Figure S3

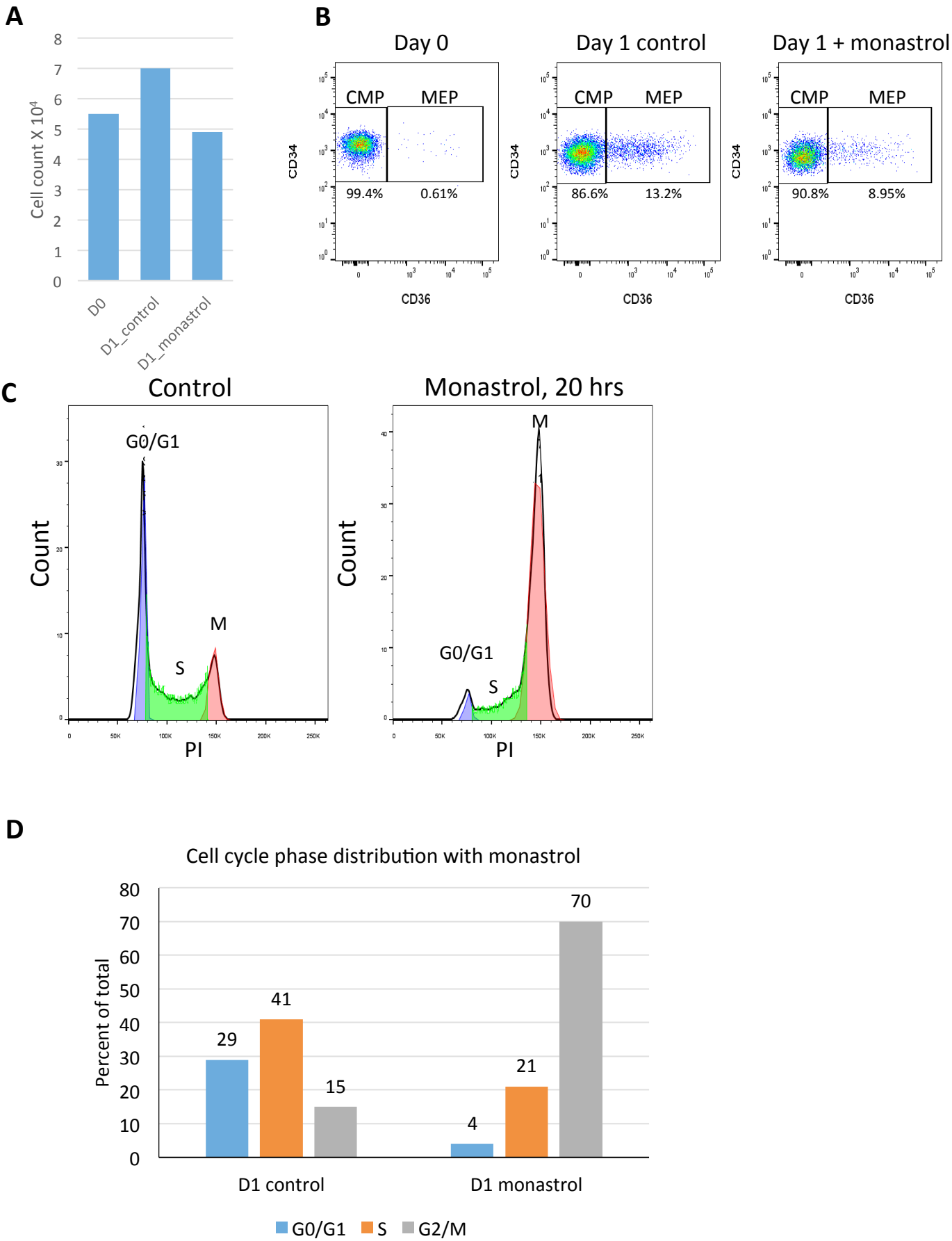

Figure S3. CMP mature to MEP independently of cell division. (A and B) $CD34^{+}CD36^{-}$ cells were sorted by FACS after 48 hours in ENM culture followed by 24 hours culture in the absence or presence of monastrol (100 μM) as indicated. **(A)** Monastrol is seen to be effective under these conditions as the cell number in the control culture increases by 1.3 fold and there is no increase in cell number in the presence of the inhibitor. **(B)** Flow cytometry analysis of cells before culture (Day 0) and after 1 day without (control) or with (+) monastrol shows that $CD36^{+}$ MEP are generated in the presence of the M phase blocker. Similar results were found with the cell cycle inhibitor aphidicolin (not shown). Results are from a single experiment done with sorted cells and are representative of results from other experiments using the same cell cycle inhibitors. **(C and D)** K562 cells were cultured without (control) or with monastrol for 20 hours and then fixed and stained with propidium iodide to determine cell cycle stage occupancy (G0/G1 phases, purple; S phase, green; M phase, pink). FlowJo vX cell cycle analysis tool was used to determine stages. 70% cells were blocked in M phase in the presence of monastrol.

Gampel, Figure S4

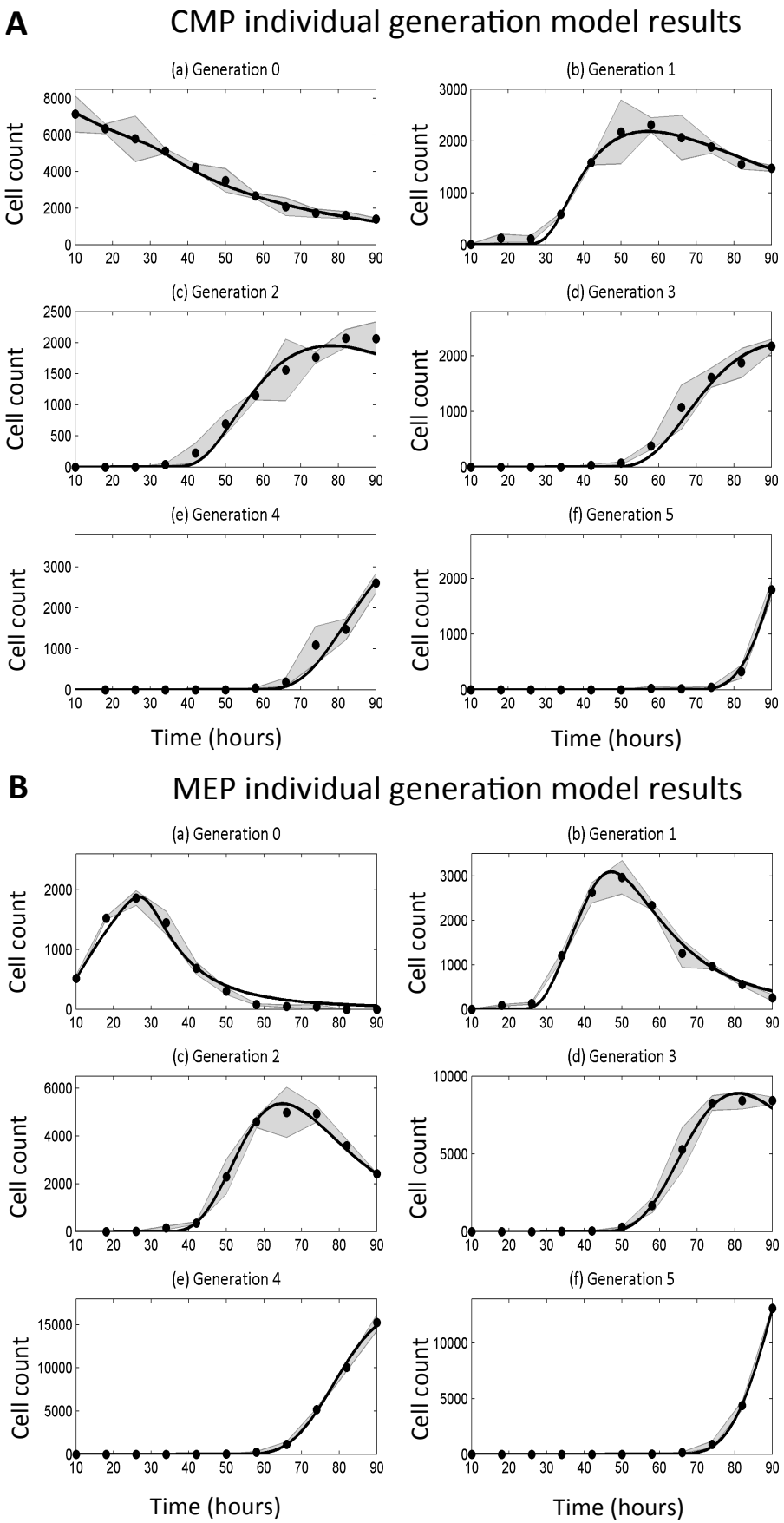

Figure S4. Individual generational data and fitted model simulations. **(A)** CMP G(0-5), model generated fits (solid line) plotted on top of experimental dataset A (circles) with the experimental standard deviation (shaded region). **(B)** MEP G(0-5), model generated fits (solid line) plotted on top of experimental dataset A (circles) with the experimental standard deviation (shaded region).

Gampel, Figure S5

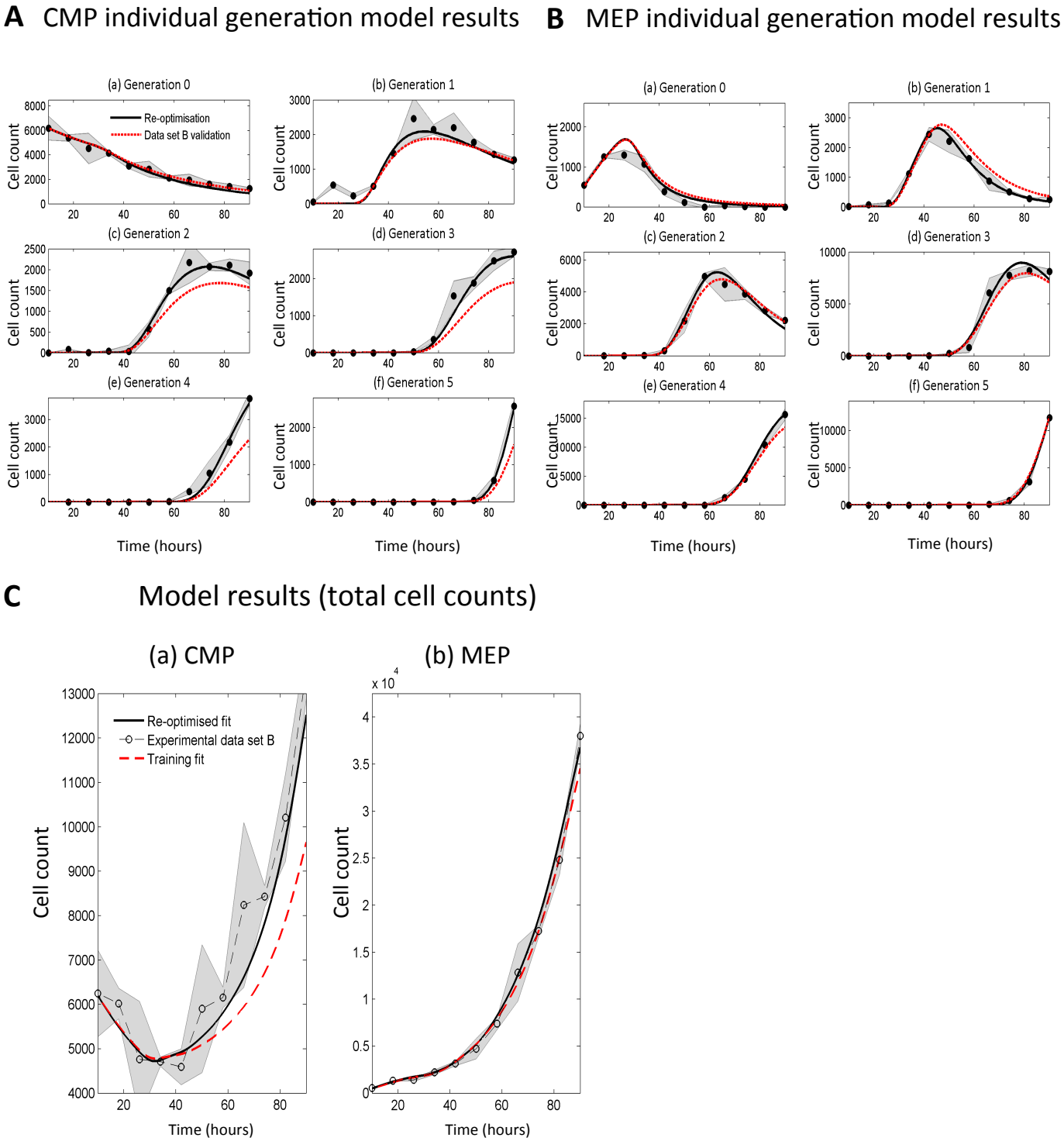

Figure S5. **Model validation and re-fitting on independent datasets. (A)** CMP G(0-5) and **(B)** MEP G(0-5) model generated predictions using parameters optimised on experimental dataset A (red dashed line) compared to model simulations using re-optimised parameters from experimental dataset B (solid black line). Experimental data is provided as dots, with standard deviation shown as shaded region. **(C)** Total cell counts of CMP (left panel) and MEP (right panel) for the validation of training parameters (red dashed line) and the re-optimised parameters (solid black line). Validation experimental dataset B shown (circles and dashed black line) with standard deviation as shaded region.

Gampel, Figure S6

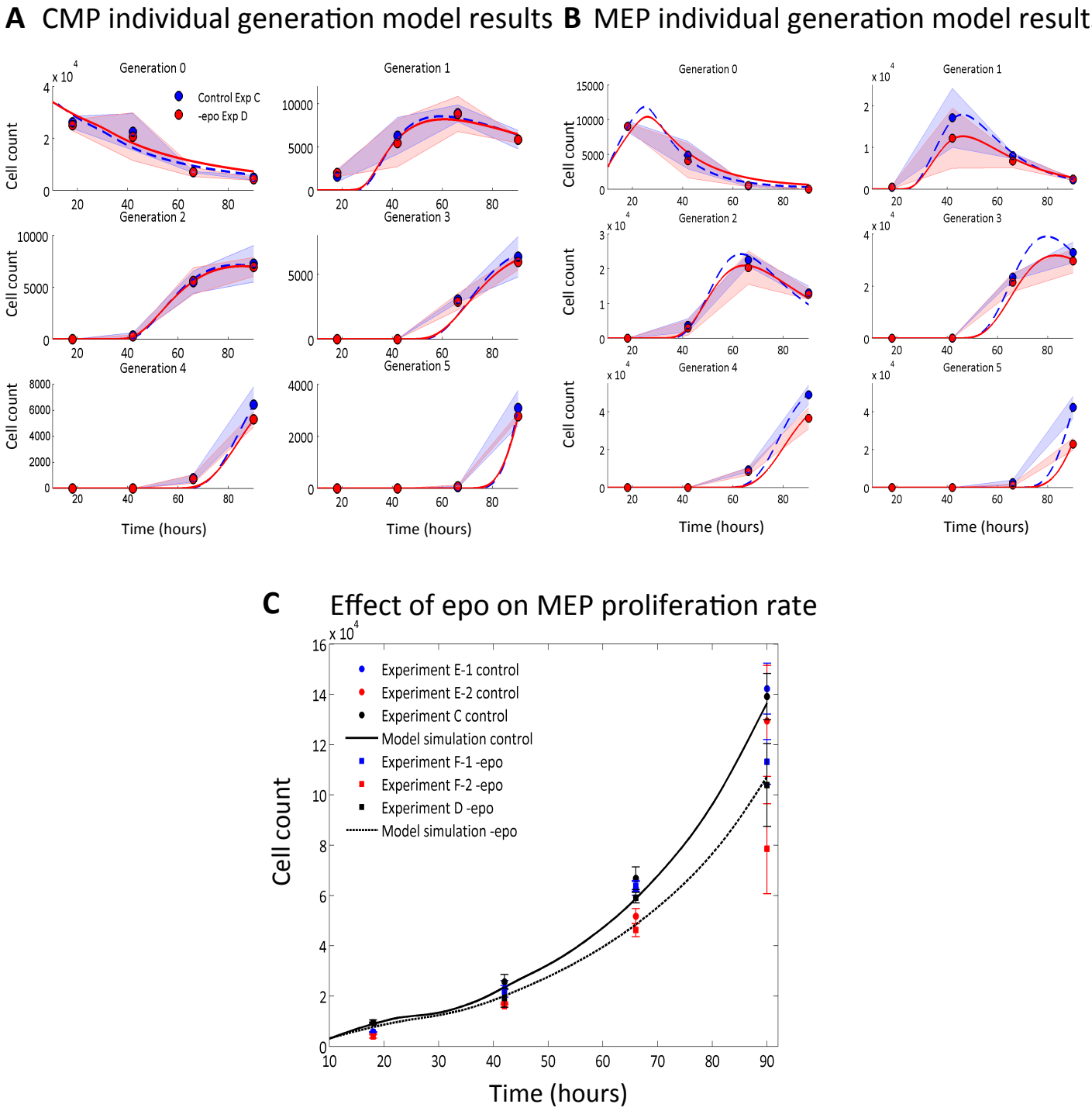

Figure S6. Individual generational fits for CMP and MEP for the control and minus epo data sets, and total MEP cell counts with validation data from two separate independent data sets. (A & B) Model fits for G(0-5) for both control dataset C (blue line) and the minus epo dataset D (red line) along with the experimental data (dots) for CMP and MEP cell populations. Standard deviation within the experimental data shown as shaded regions. **(C)** Total cell count results for MEP corresponding to the optimised parameters in Table 2 for the control and minus epo data sets. Control fit (solid black line) with experimental datasets E-1 and E-2 (dots) shown with error bars, and minus epo fit (dashed black line) with experimental datasets F-1 and F-2 (squares) shown with error bars. Validation data sets from independent experiments.

Gampel, Figure S7

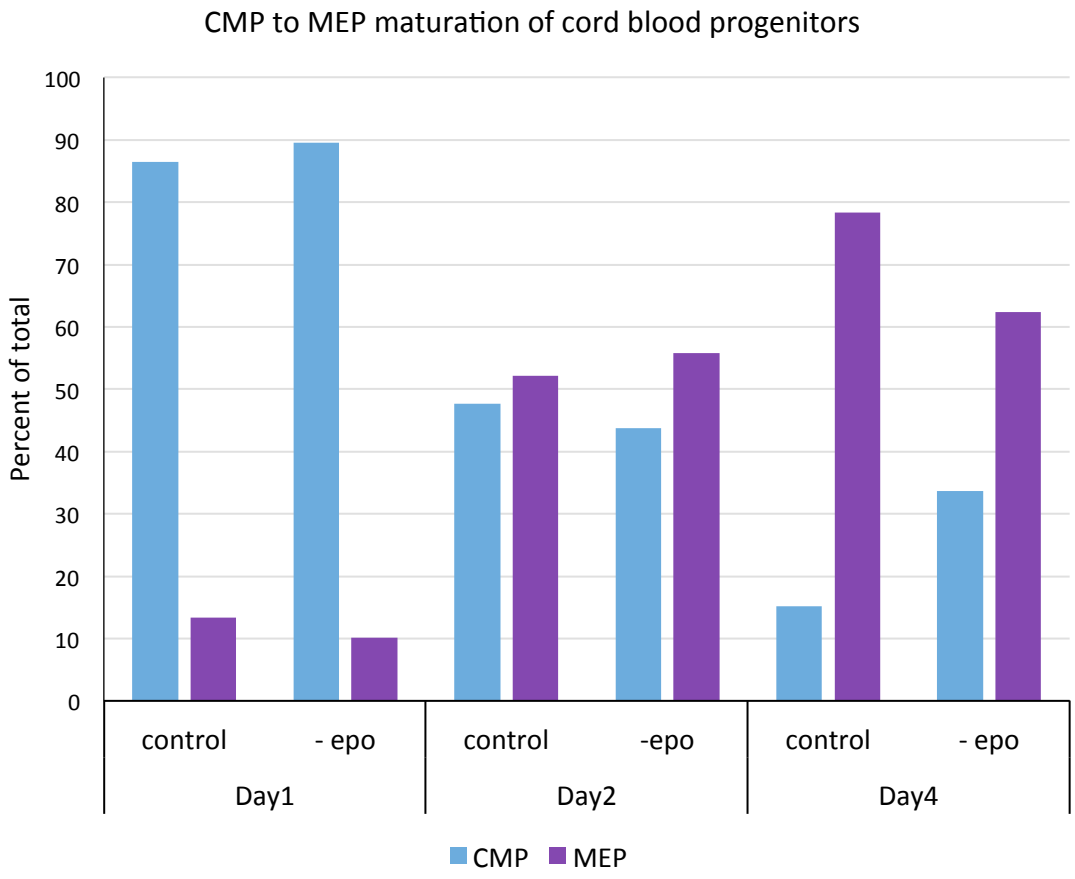

Figure S7. Population dynamics of CD34$^+$ cells from cord blood. CD34$^+$ cells from cord blood were enriched by lineage depletion and cultured in serum free medium for 4 days. Percent CMP (blue) and MEP (purple) were determined by flow cytometry gating on CD34$^+$CD36$^-$ (CMP) and CD34$^+$CD36$^+$ (MEP).