FLASH radiotherapy: Considerations for multibeam and hypofractionation dose delivery

Ranald MacKay a,*, Neil Burnet b, Matthew Lowe a,c, Bethany Rothwell c, Norman Kirkby c, Karen Kirkby c, Jolyon Hendry a

a Christie Medical Physics and Engineering, The Christie NHS Foundation Trust, Manchester; b The Christie NHS Foundation Trust, Manchester, and University of Manchester; and c Division of Cancer Sciences, University of Manchester, Manchester Cancer Research Centre, Manchester Academic Health Science Centre, UK

Current interest in ultra-high-dose-rate radiation therapy (FLASH) has arisen from the observed protection of various normal tissues exposed to high single doses in preclinical studies [1]. In this context ‘ultra-high-dose-rate’ is considered to be ≥40 Gy/s (c.f. a conventional (CONV) dose-rate ≥ 0.01 Gy/s) [2,3] (Fig. 1). While existing preclinical and clinical application of FLASH has used only a single beam, some proponents of FLASH have suggested the use of multiple beams to achieve the same conformal dose distributions obtainable using advanced radiotherapy techniques such as scanned beam proton therapy with conventional radiation dose-rates [4]. Others, including manufacturers, have also suggested the use of several high-dose FLASH fractions rather than a single dose, and even the use of lower dose fractions [5].

Recently, the clinical benefit of FLASH has been much heralded by the radiotherapy community, but FLASH is not a new phenomenon with much early pre-clinical work conducted in the 60's and 70's [6,7]. At that time, practical application was not possible due to a lack of technology to treat at FLASH dose rates. In the new age of FLASH, there are more options for delivering high dose rate but there will be important differences between the irradiation conditions in experiments that have shown a FLASH effect and the likely clinical irradiation of patients. The degree to which these differences affect the efficacy of the FLASH effect is governed by its underlying mechanism. In this work we set out some of the practical aspects of clinical irradiation and the possible consequences.

Temporal effects of FLASH delivery

The early preclinical studies by Hendry et al. [8] studied the effect of dose per pulse when irradiating mouse tails with electrons. Irradiation with pulsed 10 MeV electrons was compared with 320 kV X-rays with a half value layer of 2.3 mmCu delivered at 2.2 Gy/min. With 10 MeV electrons, the response to a wide range of dose rates was investigated by varying the dose per pulse from 0.0017 to 3 Gy, the duration of the pulse from 0.5 to 5 μs, and the total dose from 35 Gy to 70 Gy. This resulted in electron deliveries ranging between 10³ and 10⁶ Gy/s intra-pulse, with irradiation (in ambient air, awake conditions) inducing skin necrosis in mouse tails by 6 weeks as the endpoint [8]. The FLASH effect for tail skin sparing declined markedly for irradiation durations of ≥20 s (0.0017–0.055 Gy/pulse at dose rates varying from 5 Gy/min to 10,000 Gy/min), assumed due to a lower rate of O₂ depletion together with O₂ diffusion from outside the radiation field [2]. More recently Cunningham et al. [9] showed FLASH sparing when measuring the hind leg contraction in mice irradiated at dose rates of 57 Gy/s and 115 Gy/s delivered in less than a second with protons at 250 MeV. Note that where the brain was irradiated with a pulsed electron beam, a much shorter time of <200 ms was needed to observe a FLASH effect [10]. Thus there is plenty of published and emerging evidence of the FLASH effect being induced in pre-clinical irradiation.

Clinical delivery platforms will have different temporal deliveries compared to pre-clinical irradiators. Many cyclotrons employed in proton therapy have a quasi-continuous beam with pulse frequencies in MHz. Others, such as the IBA superconducting synchrocyclotron, have pulse frequencies of 1 kHz and consequently a higher dose per pulse. A recent evaluation of 127 patients treated with spot scanning proton therapy, with a higher dose per pulse, showed no reduction in toxicity despite claiming an increased dose rate of 200–1000 times faster than the classical dose rate [11]. This study offered standard treatment fractionation (1–3 Gy) to a range of benign and malignant tumours using standard scanning treatment plans. The authors quote a dose rate of ‘around 10 Gy/s per spot, depending on the range and energy needed’. Without further clarity from the publication it could be assumed that in common with all cyclotron systems the degradation of the beam to lower energies reduces the beam current and dose rate dramatically. In addition, the time to treat a volume is a combination of the dose rate of the spot, the time to scan the beam at a particular energy, and the time to change energy of the beam. The last of these, the time to change the energy of the beam, is the dominant factor for most systems. Typical volume scanning for proton therapy delivers at 2 Gy to a litre volume in 1 minute. No beam on time is quoted and, although the dose per pulse may be higher than typical, scanned beam proton therapy delivery of single beams will still be several seconds rather than sub second. In reality, the quoted increase in dose rate, to individual spots at high energy, does not reflect the dose rate in Gy/s delivered to normal tissue.
in multiple beam plans with many spots delivered at lower proton energy and consequently a reduced dose rate. Without a better understanding of the temporal effects of beam delivery on the FLASH effect, and perhaps definition of dose rate itself, extrapolating from pre-clinical to more complex clinical irradiations will be difficult.

The use of multiple ‘transmission’ FLASH beams, delivering dose at a single, high energy, rather than multi-layered ‘Bragg Peak’ treatment plans was explored in a recent paper [4]. The authors considered delivering a dose of 18 Gy per fraction using up to 10 beams to deliver dose. In the plans studied, the total irradiation time (sum of all beam times for each individual patient) varied between 317 and 730 ms. This time is not the time taken to deliver the plan to the patient, but the time that the beam is on during the delivery of a treatment fraction. For a multibeam plan the overall treatment time is the time from delivery of the first spot to delivery of the last. This includes the time it takes to irradiate the patient from each beam as well as reposition the gantry and/or patient between beams. Typical beam-on times for proton machines range from 0.5 to 5 min, though these times may be dramatically reduced by FLASH dose-rates as outlined above. When rotating, the gantry speed is limited to <6 degrees per second, i.e. <1 revolution per minute. Repositioning of the patient using the robotic couch is also comparatively slow. Even if performed remotely, the minimum time for a relatively modest rotation of the gantry or positioning of the couch and beam request to enable proton delivery from the cyclotron will be upward of a minute.

During pencil beam scanning proton therapy delivery, tissues irradiated with multiple beams may experience a range of instantaneous dose-rates from zero (between beams) to some maximum dose rate occurring at the Bragg peak of an individual pencil beam. However, any position irradiated at a high dose-rate by a single spot will also be irradiated at much lower dose rates by surrounding spots. The dose rate will be a combination of the dose rate to a given position from a number of delivered spots [12]. For a proton plan with thousands of spots, the influence of each spot on the dose rate at any given point is complex. Achieving FLASH under such circumstances may require a new optimisation approach that optimises dose rate, through the delivery of a reduced number of spots seeking to increase the proportion of dose delivered to normal tissue at FLASH dose rates. Even with such an optimisation, with multibeam delivery normal tissue doses will be well below doses where FLASH has been demonstrated.

Fractionated delivery

If a FLASH treatment were to be given as a single beam and single dose, the normal tissue response is biphasic (Fig. 2a). There is a threshold-like dose needed to induce the resistance, that is the reduction in normal tissue effects, as noted in past preclinical studies in vitro [7] and in vivo [13]. This initial region is commonly ascribed to radiochemical depletion of the available oxygen to zero (Fig. 3) [14]. It is not a true threshold with no resultant cytotoxicity, but it starts from a CONV-like response and ultra-rapidly develops into a second phase with increasing dose, due to the induced radioresistance phenomenon characteristic of anoxic cells and tissues. Thus, it is really a ‘threshold’ dose for the initiation of the FLASH-resistant second phase. This dose is not known accurately, but for target cells at low physiological O2 levels in rodent tissues, it is likely in the region of 3.5–7 Gy. The recent study on neurocognitive side-effects in irradiated mouse brain showed no difference in CONV or FLASH response using 4 fractions of 3.5 Gy, but FLASH sparing versus CONV was observed using 2×7 Gy [10]. The latter study also confirmed the lack of a FLASH effect on tumour growth restraint after all the fractionation schedules tested, which is a crucial consideration. Lower values in the threshold dose range are suggested by recent work showing that FLASH minimizes the induction of proinflammatory genes, facilitates radiation recovery and reduces the number of persistent DNA damage foci as well as cell senescence and fibrogenesis in mouse lung [15]. This deserves further investigation, hopefully including lung function assays. The radioresistant dose–response slope is about half that for CONV dose rates in the examples of rodent intestine, skin and lung [16]. The threshold and slope changes depend on the initial physoxia level in the target cells for tissue injury [17], and the beam parameter characteristics producing the FLASH effect.
FLASH radiotherapy: Considerations for multibeam and hypofractionation dose delivery

If one considers FLASH to consist of both a threshold dose and a radioresistant-phase dose, then conventional dose delivery patterns become more problematic. The 18 Gy fraction size considered by Van Marlen et al. [4], when split into ten equally weighted fields, will deliver no more than 1.8 Gy per beam. Even at FLASH dose rates this is much below the threshold dose where FLASH has been demonstrated, and hence no FLASH effect is expected. Assuming a threshold dose of 5 Gy (see below), 3 beams would be the maximum number to observe any FLASH effect and 1 or 2 beams would be preferable. Even then there is likely to be a reduction in FLASH sparing because a dose threshold will apply to each beam (Fig. 2b).

With a significant threshold dose for FLASH there are only two options for the treatment planner to induce a FLASH effect if the dose delivered by a conventional multiple beam fractionated plan delivers dose below threshold: reduce the number of beams or deliver a lower number of large fractions. It should be noted that for any normal tissue irradiated in a treatment plan, reducing the dose below threshold is the most effective way of reducing biological effect. This is true even with a strong FLASH effect. Thus when planning FLASH treatments it would be wrong to purposely increase dose to normal tissue to induce a FLASH effect without reducing the volume of tissue irradiated at conventional dose rates.

Clinical implementation and treatment concepts

The first patient prescribed FLASH received 15 Gy with 5.6 MeV electrons to a 3.5 cm diameter tumour, using ten 1 μs pulses delivered in 90 ms to a cutaneous lymphoma, which healed well [18]. A 5 mm bolus was used so that the total depth covered by the 90% isodose was 1.3 cm. Previously, such a patient would have received 20 Gy CONV as 10 fractions of 2 Gy, or 21 Gy as 6 fractions of 3.5 Gy. These schedules translate into BED\textsubscript{10} values of ~24 Gy and ~28.4 Gy respectively using the LQ formula with α/β = 10 GY for early reactions, equivalent to single doses of ~11.3 Gy and ~12.6 Gy. BED\textsubscript{3} values for late reactions are estimated to be ~33.3 Gy and ~45.5 Gy respectively, equivalent to single doses of ~8.6 Gy and ~10.3 Gy. These calculations indicate the appropriateness of the choice of dose. A caveat is that the LQ model was never intended to be used for single-dose extrapolations, and hence the calculated values are only rough estimates.

However, it must be re-emphasised that the BED concept cannot be used directly for calculations using fractionated FLASH exposures [16], as done recently for brain experiments [10], if a normal tissue sparing effect actually exists. The BED concept was invented to produce a metric which could quantify the relative strength of different schedules of conventional dose-rate fractions, typically of magnitude 1–6 Gy, for early and late adverse events and tumour responses [19]. Extra factors can be incorporated to accommodate (a) incomplete repair due to multiple fractions per day or short exposures at low dose rate, and (b) longer overall treatment times due to repopulation. These are all sensitising factors, reducing the effective BED. In contrast, FLASH produces radioresistance (to normal tissue damage) at doses above a ‘threshold’ dose. This requires an extra ‘resistance factor’ (RF) to express FLASH doses in a form compatible with BED values for CONV irradiations (Fig. 2).

FLASH-equivalent doses can be expressed in two ways. Firstly the FLASH equivalent dose to a CONV irradiation, can be calculated, assuming a FLASH sparing effect on normal tissues. This amounts to a higher dose than the nominal prescription dose, and is achieved by multiplying the dose delivered under FLASH conditions by the RF. Secondly, the dose ‘seen’ by normal tissue and representative of the expected toxicity for CONV irradiations can be calculated, which is smaller than the nominal prescription dose, following division of the dose delivered by FLASH by the RF. When considering a threshold above which FLASH occurs the RF is applied to dose above threshold.

If the clinically delivered 15 Gy FLASH dose was comprised of, say, a threshold dose ~5 Gy and a dose of ~10 Gy in the FLASH-resistance mode using the preclinical CONV/FLASH slope ratio (or RF) of ~1.7 measured for rat foot skin [11]. The actual expected toxicity would be equivalent to a CONV dose of (~5 + 10)/1.7 = 1 0.9 Gy, using the same preclinical RF of ~1.7. The tolerance single dose for human skin late reactions using a 2.0–3.0 cm diameter field is ~20 Gy CONV [20]. This was based on Christie Hospital (Manchester, UK) data showing a skin necrosis rate of 2.3%, 3.0% and 8.9 % by 7 years after 18 Gy, 20 Gy and 22.5 Gy single dose CONV respectively, for the treatment of skin cancer in a series of

![Fig. 2](image-url)
patients. Hence the previous CONV doses used for skin cancer and cutaneous lymphomas are consistent with the dose chosen for the first FLASH treatment.

Note that the Dose Modifying Factor (DMF) of FLASH is a clinically useful term for prescribing isoeffective doses. This is the ratio of doses of FLASH and CONV producing the same endpoint, which incorporates both threshold and dose–response slope components. If the threshold dose was zero, the DMF would be simply the CONV/FLASH slope-ratio. When the FLASH threshold dose is positive, that dose would need to be subtracted from both the CONV and the predicted FLASH dose to produce respective dose components. The initial CONV-like component would be multiplied by the CONV/FLASH slope-ratio, or RF, to produce the radioresistant dose component, added to the FLASH threshold dose to calculate the total FLASH dose, and then divided by the total isoeffective CONV dose to give the DMF. This would be a first approximation. A second approximation could also take into account the gradation of dose–response at the inflexion region between the CONV and FLASH components, which could be suitably modelled. This correction would be slightly larger at higher doses showing greater FLASH effects. DMF values of 1.2–1.3 have been discussed, arising from preclinical studies on cats and pig skin [21]. These lower values could represent greater heterogeneity in responses producing lower average values of parameters.

In the prescription planning for the first patient [18], the dose to normal tissue was considered likely to be about 2/3 of the nominal prescription dose, which in effect uses a DMF of 1.5 (i.e. the reciprocal of 2/3). Thus, 2/3 of 15 Gy would amount to ~10 Gy, close to the slightly more detailed calculation above, suggesting 10.9 Gy as equivalent.

Where doses are fractionated, like the example of 3×18 Gy for lung tumours [4], the FLASH effect will be diluted because of the presence of the threshold dose as a component part of each treatment fraction. Using example numbers for protons, if the threshold dose in each fraction is, say, 5 Gy and the induced resistance (RF) for higher doses is 1.5 to 2, the FLASH toxicity dose equivalent would be approximately between $5+(13 \times 1.5)$ and $5+(13 \times 2) = 24.5$ to 31 Gy. If the threshold dose is lower, the equivalent doses would be slightly higher. The effective maximum dose gain or DMF from a full FLASH effect in this example would be $(24.5 \div 31)/18 = 1.4$ to 1.7 fold, remarkable elevations of the FLASH effect dose in normal tissues irradiated to the prescription dose. However, as noted above, in a multibeam plan areas irradiated by one beam would not be irradiated to threshold and thus would not be spared by FLASH.

The abolition of the FLASH effect by the use of physioxia modifiers in the above two preclinical rodent studies (skin temperature increase, carbogen breathing for brain) flag up a warning for FLASH proponents. If effective increases in tolerance dose were planned using the FLASH effect, enabling higher levels of tumour control, it would be essential to have an in vivo test of whether the FLASH effect had happened or not and by how much. If such a test was available, it would be better to use 2 beams or 2 fractions, so that the result of the first dose could guide the delivery/size of the second. A method has been described recently and is still under development, based on the oxygen–dependent phosphorescence lifetime of the probe Oxyphor 2P in vivo [22]. If this could be coupled to stem cell markers it may provide an even greater specificity of oxygen depletion and induced radioresistance in target regenerative cells in FLASH-irradiated tissues.

**Extension of clinical studies required to demonstrate normal tissue sparing**

The safe, ethical development of clinical studies to provide further human data on both normal tissue sparing and tumour efficacy is challenging. The first imperative is to ensure safety by demonstrating reproducible, inter-patient sparing of normal tissue response. It might be possible to find patients with metastatic disease in whom two different metastases could be treated differently, one with CONV and the other with FLASH, imaging biomarkers of normal tissue response [23] could be used to compare the effects of the two approaches, for example in lung, although the endpoint read-outs develop long after completion of treatment.

Loss of gas exchange capability in a narrow corridor of lung from such a strategy (or 2 corridors if 2 metastases were treated with different approaches) may be safe for the patient. However, considerations in the brain are less simple. If a metastasis was treated using transmission proton-FLASH, a cylinder of normal brain
tissue would be at risk of toxicity. Disregarding acute effects, risk of necrosis can be estimated using the LQ formulation and methods above. Since a typical stereotactic radiosurgery dose would be 24 Gy (in a single fraction) [24], taking advantage of the volume effect, Table 1 shows the challenge of achieving safety while delivering a (tumour) therapeutic dose, depending on the FLASH normal tissue sparing (RF) which can be achieved clinically.

Proton-FLASH is currently proceeding using high energy transmission beams which loses one of the key advantages of proton beam therapy, namely the Bragg peak. The consequent increase in dose to normal tissue will need to be balanced by the reduction in effect due to FLASH. This will have to be assessed once more is known of the clinical physiological responses.

The considerations of gantry move time, threshold dose per fraction, and FLASH effect uncertainty lead to a conservative conclusion that either 2 large FLASH fractions or a single big boost dose plus CONV as suggested earlier [2] would be reasonable delivery strategies. An in vivo FLASH-effect dosimeter would be valuable to check delivery and allow any compensation for a second dose. A time between fractions of perhaps greater than several hours may allow physiological conditions to return close to the original status.

Conclusions

The current interest in ultra-high-dose-rate FLASH is based on recent observations of sparing of various normal tissues in preclinical studies [1]. Older evidence of the FLASH phenomenon also exists to support this. It is known that we can induce the FLASH effect in partially hypoxic normal tissues. The induction of the FLASH effect is dependent on these normal tissues being irradiated at high dose rate and there is evidence to suggest there may be a dose threshold below which the FLASH effect is not seen. The recent preclinical data suggest equal tumour control but this will need to be confirmed. The mechanism behind FLASH remains unclear, and a better understanding of this will help to realise the full potential of this approach.

FLASH is fundamentally a property of irradiated tissue and if the FLASH effect is to be evoked during radiotherapy treatment in tissues that we normally seek to spare from irradiation then there are many questions to be answered. The irradiation of normal tissue in an advanced radiotherapy technique such as protons is more complex than the irradiation in the pre-clinical setting. In proton therapy there is often no such single dose rate delivered to normal tissue as any point irradiated will receive its dose from a number of proton spots. The relation between the dose rate of delivery and the dose rate to tissue where we wish to induce FLASH may not be simple. Before we can confidently utilise FLASH in the clinic there are a number of issues that require clarity:

1. More information is required on dose and dose-rate thresholds, and on possible variation between tissues and between individuals.
2. Solutions to the challenges of using multiple beams (physical) and fractions (biological) to spare normal tissues when using FLASH dose rates are needed.
3. If an initial dose is needed before the FLASH effect begins, this needs to be factored into calculations of FLASH-equivalent doses and into treatment planning.

These factors are essential in considering the trade-off between FLASH and the use of standard fractionation to achieve normal tissue sparing. Clinical studies are needed first to demonstrate normal tissue sparing and then efficacy; constructing such studies remains a challenge.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Professor Karen Kirkby and Professor Ranald MacKay are members of the Varian Flash Forward consortium and The University of Manchester holds and conducts research grants concerning Flash for which Professor Kirkby is the principal investigator.

Acknowledgements

This work was supported by Cancer Research UK via funding to the Cancer Research Manchester Centre, Research UK RadNet Manchester and the Advanced Radiotherapy Technologies Network (ART-NET). Professor Karen Kirkby and Professor Ranald MacKay were supported by NIHR Manchester Biomedical Research Centre. This work was supported by the European Unions H2020 Research and Innovation Programme under Grant Agreement No. 730983.

References


