Hypoxia-activated prodrugs and (lack of) clinical progress: The need for hypoxia-based biomarker patient selection in phase III clinical trials

Linda Spiegelberg a, Ruud Houben b, Raymon Niemans a, Dirk de Ruyscher b, Ala Yaromina a, Jan Theys a, Christopher P. Guise c,d, Jeffrey B. Smaill c,d, Adam V. Patterson c,d, Philippe Lambin a, Ludwig J. Dubois a,⇑

a Department of Precision Medicine, The M-Lab, GROW – School for Oncology and Developmental Biology, Maastricht Comprehensive Cancer Centre, Maastricht University Medical Centre, Maastricht, The Netherlands
b Department of Radiation Oncology (MAASTRO), GROW – School for Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht, The Netherlands
c Translational Therapeutics Team, Auckland Cancer Society Research Centre, School of Medical Sciences, University of Auckland, Auckland, New Zealand
d Maurice Wilkins Centre for Molecular Biodiscovery, School of Biological Sciences, University of Auckland, Auckland, New Zealand

A R T I C L E   I N F O

Article history:
Received 7 December 2018
Revised 11 January 2019
Accepted 13 January 2019
Available online 18 January 2019

Keywords:
Hypoxia-activated prodrugs
Phase III clinical trial
Biomarker
Hypoxia

A B S T R A C T

Hypoxia-activated prodrugs (HAPs) are designed to specifically target the hypoxic cells of tumors, which are an important cause of treatment resistance to conventional therapies. Despite promising preclinical and clinical phase I and II results, the most important of which are described in this review, the implementation of hypoxia-activated prodrugs in the clinic has, so far, not been successful. The lack of stratification of patients based on tumor hypoxia status, which can vary widely, is sufficient to account for the failure of phase III trials. To fully exploit the potential of hypoxia-activated prodrugs, hypoxia stratification of patients is needed. Here, we propose a biomarker-stratified enriched Phase III study design in which only biomarker-positive (i.e. hypoxia-positive) patients are randomized between standard treatment and the combination of standard treatment with a hypoxia-activated prodrug. This implies the necessity of a Phase II study in which the biomarker or a combination of biomarkers will be evaluated. The total number of patients needed for both clinical studies will be far lower than in currently used randomize-all designs. In addition, we elaborate on the improvements in HAP design that are feasible to increase the treatment success rates.

© 2019 The Authors. Published by Elsevier B.V. on behalf of European Society for Radiotherapy and Oncology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Tumor hypoxia is a well-known tumor microenvironmental parameter present in most solid tumors, which hampers the efficacy of conventional anti-cancer treatments. Blood vessels within rapidly expanding tumor tissue often fail to develop properly, being primitive (dilated and leaky), chaotic (irregular and tortuous) and dysfunctional (blind ends and arteriovenous shunts). Two forms of tumor hypoxia can be distinguished, namely diffusion-limited (chronic) and perfusion-limited (acute) [1]. Both radiotherapy (RT) and chemotherapy are dependent on the blood supply to exert their effects. In the case of radiotherapy, oxygen reacts rapidly to modify the lesion that is caused by ionizing radiation producing permanent DNA damage [2]. In the absence of oxygen much of the damage can be repaired by the cells themselves, rendering hypoxic cells three times more resistant to radiation [3]. Chemotherapeutic resistance is caused by several hypoxia-related factors. First of all, the hypoxic cells within the tumors are difficult to reach, existing in a pharmacological sanctuary due to the aberrant blood supply. Additionally, decreased cellular proliferation, lost sensitivity to p53-mediated apoptosis and upregulation of genes involved in drug resistance also contribute to hypoxia-related chemoresistance [4,5]. Furthermore tumor hypoxia leads, through hypoxia inducible factor (HIF)-related gene expression and the unfolded protein response (UPR), to an increased metastatic potential and thus worse outcome [6]. Severe hypoxia is an attractive target for anti-cancer therapies [7], since it is uniquely present in tumors and is a key factor that leads to rapid disease progression and poor prognosis [8].

Tumor hypoxia can be addressed in different ways and approaches are primarily based on oxygen modification strategies, oxygen mimetics and cytotoxic agents. Oxygen modification strategies aim to either increase tumor oxygenation or decrease oxygen consumption of cells. Hyperbaric oxygen therapy, hyperthermia and carbogen breathing combined with nicotinamide have been
used in clinical trials as adjuvant therapies to increase tumor oxygenation. In hyperbaric oxygen therapy, 100% oxygen is inhaled under elevated pressure, leading to systemically increased oxygen tension [9]. Hyperthermia, the mild local elevation of temperature, leads to dilatation of blood vessels thereby stimulating blood flow [10]. Carbogen (95% oxygen, 5% carbon dioxide) breathing in combination with the vasodilator nicotinamide also increases blood flow, thereby decreasing hypoxia. The latter has been clinically investigated in combination with accelerated radiotherapy (ARCON trial) [11,12]. However, the beneficial effect on survival and outcome of these therapies is still debated and the high costs, difficulty of practical planning and toxicities prevent them from wide clinical use [13–16]. Decreasing hypoxia by reducing the cellular oxygen consumption using e.g. metformin, an inhibitor of mitochondrial complex 1 activity [17], has been shown to increase tumor radiosensitivity in a mouse xenograft model [18] and is currently under investigation in a Phase II trial in cervix cancer (NCT02394652). Oxygen mimetics are used to sensitize hypoxic cells to radiation by replacing oxygen in the millisecond chemical reactions needed to fix DNA damage. Although in vitro and in vivo studies were promising, clinical use was hampered by the high doses that were needed to achieve radiosensitizing effects, giving rise to significant toxicities [19]. Lastly, hypoxic cells can be directly sterilized by hypoxia-activated prodrugs (HAPs). Different classes of HAPs exist, all of which are activated by reduction facilitated by cellular oxidoreductases [20]. Typically, the initial reduction event is reversible in the presence of oxygen. Under hypoxia, DNA-reactive cytotoxins are formed that kill the hypoxic cells [21]. Several HAPs, as recently has been reviewed [22], have been developed and are under extensive preclinical and clinical evaluation. This review summarizes the (pre)clinical development of three of the most clinically advanced HAPs. Tirapazamine (TPZ), PR-104 and TH-302, address different hypoxia-related biomarkers and finally proposes a clinical trial design with biomarker assessment for phase III studies that may result in positive trials and thus clinical implementation of this promising anti-cancer therapy.

2. (Pre)clinical development of HAPs

2.1. Tirapazamine

TPZ is the first HAP that was evaluated and after extensive preclinical testing, demonstrated clinical safety in 1994 [23]. Despite early promise, results of phase III clinical trials were disappointing; no therapeutic benefit could be established compared to standard chemoradiotherapy or chemotherapy alone [24]. However, the outcome of the trial was hampered by the poor quality of radiotherapy and if corrected for this, the TPZ treatment arms did perform better. It was further hypothesized that the lack of effect was due to excessive drug consumption leading to poor extravascular transport, coupled with the observation that TPZ is activated under relatively mild hypoxia also present in liver, gastrointestinal (GI) tract and bone marrow, which may have contributed to toxicities. Notably, the aromatic N-oxide class is prone to rapid redox cycling [25], which may account for other dose-limiting side effects such as muscle cramping and severe fatigue. Discrepancies between murine and human tolerance are an important cause of cell kill, with AKR1C3 metabolism negating hypoxia targeting and expression was also shown in normal human tissues [38]. Notably, AKR1C3 has also been reported to be expressed in myeloid cell lineages where it has been proposed to play an important role in regulating cell proliferation and differentiation [39,40]. The expression of AKR1C3 could, at least in part, explain the dose-limiting bone marrow toxicities observed for PR-104 in the clinic and likely contributed to the failure to dose-escalate PR-104 to therapeutic levels. Two of the above mentioned clinical studies that were terminated [36] and NCT00862134 were based on the high AKR1C3 expression in the tumor (hepatocellular carcinoma and non-small cell lung cancer respectively). The possibility to use tumor AKR1C3 expression as an individualized target for PR-104 treatment has not been proven successful, and the activation of the prodrug in normal tissues opposes its further use in anti-cancer treatment.

2.2. PR-104

Studies have shown that the activation of PR-104A is more readily inhibited by oxygen than TPZ, with a reported K-value (oxygen concentration to halve cytotoxic potency) which is 10-fold lower than that of TPZ [28]. Activation gives rise to a relatively stable cytotoxic metabolite that can diffuse from the hypoxic cell to neighboring, well oxygenated cells, creating a localized bystander effect [29]. PR-104, a phosphate pre-prodrug of PR-104A, is a dinitrobenzamide mustard that has undergone broad preclinical and clinical investigation. The HAP contains a latent nitrogen mustard moiety that becomes activated under severe hypoxia and causes cell kill by inducing DNA cross-links [30]. Several preclinical studies have investigated the effect of PR-104 in different xenograft tumor models, either as monotherapy or in combination with standard anti-tumor therapies. In in vivo therapeutic studies PR-104 was proven to be effective, in terms of increased cell kill, inhibition of tumor growth or increased mouse survival. Furthermore, combination of PR-104 with radiotherapy or chemotherapies enhanced these effects [28,30–33]. Direct comparison with TPZ indicated superiority of PR-104 presumably caused by its bystander effect [28,30]. This bystander effect is hard to prove in vivo, but in vitro models have been established. Wilson and colleagues used a multicellular layered cell culture system and showed a lack of bystander effect for TPZ, whereas three dinitrobenzamide agents provided efficient bystander effects [29].

Clinical safety and tolerability of PR-104 was evaluated in patients with solid tumors refractory to standard treatment [34]. In this study, with an every 3-week schedule, PR-104 was well tolerated with neutropenia as the primary toxicity. McKee and colleagues investigated a weekly administration schedule and found that thrombocytopenia (decrease of thrombocytes leading to excessive bleeding) and neutropenia were the dose limiting toxicities (DLTs). Therefore, a short course of treatment combined with radiotherapy was proposed [35]. However, the phase I study of Abou-Alfa and colleagues, in which PR-104 was combined with the tyrosine kinase inhibitor sorafenib in advanced hepatocellular carcinoma, was stopped because the therapy was poorly tolerated by patients [36]. Combination with the chemotherapeutics gemcitabine or docetaxel in advanced solid tumors was also halted due to dose-limiting thrombocytopenia [37]. Another study using PR-104 and docetaxel in non-small cell lung cancer was terminated early because interim analysis showed a low probability of significant results (NCT00862134).

Apart from the hypoxic activation of PR-104, Guise and colleagues showed that aldo-keto reductase 1C3 (AKR1C3) is able to reduce PR-104 into its active form independent of oxygen [38]. AKR1C3 is highly expressed in different human tumor cell lines and could therefore provide a more individualized target for PR-104 treatment of patients. However, AKR1C3 metabolism negates hypoxia targeting and expression was also shown in normal human tissues [38]. Notably, AKR1C3 has also been reported to be expressed in myeloid cell lineages where it has been proposed to play an important role in regulating cell proliferation and differentiation [39,40]. The expression of AKR1C3 could, at least in part, explain the dose-limiting bone marrow toxicities observed for PR-104 in the clinic and likely contributed to the failure to dose-escalate PR-104 to therapeutic levels. Two of the above mentioned clinical studies that were terminated (NCT00862134) were based on the high AKR1C3 expression in the tumor (hepatocellular carcinoma and non-small cell lung cancer respectively). The possibility to use tumor AKR1C3 expression as an individualized target for PR-104 treatment has not been proven successful, and the activation of the prodrug in normal tissues opposes its further use in anti-cancer treatment.

2.3. TH-302 (Evofofosamide)

TH-302 is a 2-nitroimidazole-based nitrogen mustard prodrug that is reduced under hypoxia, leading to the release of isophos-
phoramide mustard (IPM), which alkylates DNA. However, the IPM active metabolite is charged at physiological pH suggesting it will not readily diffuse through membranes to surrounding cells to exert its cytotoxic effect [41]. Preclinical research regarding the effect of TH-302 in in vivo tumor models is more widespread. As for PR-104, TH-302 has been tested as monotherapy as well as in different combinations with existing anticancer therapies. TH-302 monotherapy inhibited tumor growth or prolonged survival and combination with cheo- or radiotherapy significantly enhanced the effect in most studies and models [42–44]. In an extensive monotherapy study using 8 xenograft models, Sun and colleagues found a good correlation between the hypoxic fraction of the tumor measured at baseline by pimonidazole immunohistochemical staining and TH-302 efficacy [45]. In the three non-responding tumor models, hypoxic fractions were below 5%. Additionally a causal relationship was found between the tumor oxygenation levels and the therapeutic efficacy. This effect was confirmed in a study of Peeters and colleagues [46] and provides a proof-of-principle for TH-302 activity in hypoxic tumor cells. Furthermore, several studies have shown that the response to TH-302 could be predicted using different imaging methods [46,47]. Besides the combination of TH-302 with cheo- or radiotherapy, other approaches have been used to augment the effect of TH-302. The exacerbation of transient hypoxia by either hydralazine or pyruvate has been tested and in both cases tumor growth delay was increased [48–50]. Hypoxia modification seems therefore feasible and triple combination with cheo- or radiotherapy could further enhance treatment outcome. However, optimal treatment schedules should be carefully considered since the increased hypoxia may oppose the effect of cheo- or radiotherapy. Other recent studies aimed to enhance the cytotoxicity of TH-302 by sensitizing tumor cells to DNA-damage induced apoptosis using Chk1 or mTOR-inhibitors and showed enhanced anti-tumor activity [51,52].

Clinical safety and therapeutic efficacy testing in several Phase I and II studies with TH-302 led to promising results [53–56]. TH-302 was well tolerated with dose-limiting skin and mucosal toxicities. The combination of TH-302 with doxorubicin resulted in increased hematologic toxicity of doxorubicin, but this was manageable with prophylactic growth factor support. Evidence of anti-tumor activity was established, as well as a favorable progression-free survival, overall survival and tumor response. This paved the way for Phase III clinical trials, the results of which were eagerly awaited. Two extensive trials, with more than 600 patients each, were carried out in advanced pancreatic cancer (MAESTRO; NCT01746979) and soft tissue sarcoma (TH CR-406/SARC021) and evaluated the effect of the addition of TH-302 to conventional therapies (gemcitabine and doxorubicin, respectively) on overall survival [57]. Both trials failed to meet their primary endpoints of improved overall survival. However, for the MAESTRO trial, the hazard ratio of 0.84 nearly reached significance (p = 0.0588). The researchers pointed out three potential factors explaining these disappointing findings. Firstly, the placebo-group performed better than the initial assumptions. Secondly, slightly more patients in the control-arm received second-line therapy following disease progression than in the experimental arm [58]. Finally, intent-to-treat rules led to 2 patients out of 693 being randomly assigned to receive TH-302 + gemcitabine but, due to delays, eventually led to receiving gemcitabine + placebo following re-randomization. Statistical analysis of overall survival by treatment-received, rather than intent-to-treat, did achieve significance (p = 0.0485).

Because of the negative results in these trials, TH-302 appears to follow in the footsteps of TPZ, whereby neither have achieved positive Phase III results after promising Phase I and II studies. Remarkably, all trial designs lacked one critical feature, namely the assessment of the levels of tumoral hypoxia. It has been shown in a broad range of tumors that hypoxia levels can vary widely [59,60]. For pancreatic cancer, values ranging from 0 to 26% are not reported [61]. Patients with a low hypoxic fraction are not expected to benefit from combination treatment. In the MAESTRO trial, where results were approaching significance, it would have been of great importance to have information regarding the hypoxic status of the tumors of each individual patient.

### 3. Improved hypoxia-activated prodrugs

Each HAP candidate is a bespoke invention [42,43,62]. The diversity of pharmacophores and their mechanism of action indicate every HAP candidate will have tailored requirements as design criteria for optimal activity are stringent [63]. It is notable that several HAPs were identified through in vitro screening campaigns that selected for the pharmacodynamic (PD) endpoint of maximal individual cell kill under low cell density conditions [64,65]. This process of employing anti-proliferative assays generally favors selection of HAP candidates with high rates of reductive metabolism coupled with poor or zero diffusion of cytotoxic metabolites, since both features act to maximize the measured endpoint of individual cell sterilization in low-cell density monolayers. It has been demonstrated that selection of an optimal HAP benefits from sophisticated multi-parameter modeling to carefully balance drug diffusion/consumption for adequate tissue penetration and thus maximize distal hypoxic cell kill [28,66,67]. For example, the clinical failure of TPZ is likely due, in part, to poor tissue penetration [66,67] and inadequate oxygen inhibition ($K_{O2} = 1.3 \pm 0.28 (\mu L)$) [28], with toxicity preventing schedule/dose intensification [68], HAPs such as PR-104 [30] possess several optimal properties, including good extravascular transport (tissue penetration) [21,63], strict oxygen inhibition ($K_{O2} = \sim 0.126 \pm 0.021 (\mu L)$) [28] and adequate cytotoxic metabolite redistribution ('bystander effect') [29,69]. However, several unforeseen problems led to the subsequent clinical failure of PR-104, most notably the aerobic activation of PR-104A by human aldo-keto reductase 1C3 (AKR1C3) [38] and high levels of circulating cytotoxic metabolites [70]. Both features likely contributed to the dose-limiting myelotoxicity in clinical trials [35].

Optimizing HAP design for maximal bystander effect is also challenging and typically encourages selection of candidates with more lipophilic metabolites, ignoring the rules of lipophilic efficiency which can have negative consequences, such as high protein binding, excessive microsomal metabolism and poor formulation properties [71–73]. Further, given that bystander effects operate at the micron scale (<0.1 mm) and intratumor heterogeneity of hypoxia is generally a macro scale (>10 mm) phenomenon, the proposed solution does not strictly address the problem. Thus, while controlled metabolite redistribution may exert certain benefits such as overcoming localized cell-to-cell heterogeneity of oxidoreductases or oxygen concentration, it is not the panacea of successful HAP design. There is a need for predictive biomarkers to guide clinical development of HAPs, including identification of the oxidoreductase enzymes necessary to catalyze their activation via electron donation. The human flavoproteome, comprising 79 unique flavoenzymes [74], likely plays a major role in the bioreductive transformation of HAPs, which is in agreement with the known involvement of individual oxidoreductases [75–89]. Approaches aimed at identifying these key catalytic proteins, their relative contributions and tissue distributions will ultimately guide the clinical application of HAPs.
4. Hypoxia selection by biomarkers

In different studies, tumor hypoxia was shown to be a prognostic biomarker, indicative for treatment outcome independent of the applied therapy [90,91]. On the other hand, it could also be applied as a predictive biomarker, potentially forecasting the efficacy of treatment. The need to measure hypoxia is therefore evident and, to date, multiple approaches exist to detect hypoxia either directly or indirectly.

Direct pO2 measurements using oxygen electrodes inserted into the tumor have been used extensively in research articles to determine the oxygenation status of solid tumors. The procedure is safe, although highly invasive and thus repeated measurements are not feasible. Furthermore, since no discrimination between necrotic and hypoxic areas can be made, the amount of hypoxic tissue can be overestimated and the fact that it requires skillful personnel to operate the system makes the inter-operator variability high [92]. Direct imaging of oxygen using 18F MRI or blood oxygen level dependent MRI (BOLD MRI) is also possible, however with their own limitations. In the case of 133P MRI, local injections into the tumor are necessary while in BOLD MRI the signal can be influenced by factors other than hypoxia, resulting in low specificity [93]. A promising MRI technique proposed by O’Connor and colleagues is oxygen enhanced MRI (OE-MRI), which is less invasive and potentially more specific [94,95].

Indirect methods can be based on exogenous or endogenous markers for hypoxia. Exogenous injectable markers include different 2-nitroimidazole compounds such as pimonidazole and EF5 [90]. They form stable adducts with macromolecules only at low oxygen tension, which can be detected by specific antibodies and quantified semi-automatically [96]. However, additional tumor biopsies and expertise in staining quantification are needed for this purpose. Therefore clinical usage remains limited and validation is still needed. When labeled with 18F, these 2-nitroimidazole compounds can also be used to image hypoxia using noninvasive positron emission tomography (PET). Well-known PET tracers for hypoxia include FMISO, FAZA and HX4 (extensively summarized in [97–99]). FMISO, the first tracer that was developed and the one that has been studied most extensively, could identify hypoxia in different human tumors [60], although limited clearance of the unbound tracer due to its high lipophilicity leads to low tumor specificity. FAZA and HX4 partly overcome that problem because they are more hydrophilic, FMISO and FAZA were clinically shown to have prognostic potential [100,101]. In a simulation study, HX4 showed the highest clearance and image contrast, but also the largest patient-to-patient variability [102]. A preclinical study that compared HX4, FAZA and FMISO in a rat rhabdomyosarcoma model showed different characteristics of these tracers regarding tumor-to-background ratio, spatial reproducibility and sensitivity to oxygen modification, perhaps making it a challenge to identify the optimal hypoxia tracer [103]. Endogenous markers of hypoxia are based on the biological consequences of hypoxia. Hypoxia inducible factor 1 (HIF-1) is stabilized under hypoxic conditions, which in turn regulates the expression of certain proteins and genes. In this respect the expression of HIF-1 itself, and its downstream targets such as carbonic anhydrate IX (CAIX) and glucose transporter 1 (Glut-1) have been investigated immunohistochemically and it was shown that elevated expression is in general associated with poorer outcome in patients with certain solid tumors [104,105]. However, correlation with direct pO2 measurements is minimal and these proteins can be influenced by factors other than hypoxia.

To improve specificity, various hypoxia gene expression signatures have been developed by different groups [106]. For example, Toustrup and colleagues [107] identified 15 hypoxia responsive genes in head and neck squamous cell carcinomas that could characterize the hypoxic state of a tumor and showed that they were associated with a worse clinical outcome. Also, this outcome could be improved by hypoxia modification using nimorazole. The measurement of secreted markers in the blood would be a faster and easier method to establish tumor hypoxia, without the need for biopsy material. In this respect, plasma osteopontin (OPN) has been shown to be associated with tumor pO2 in a few studies [104]. Furthermore, OPN levels were shown to be an independent prognostic marker for head and neck squamous cell carcinoma (HNSCC) [108,109]. Although this method is non-invasive and offers the opportunity to do serial measurements, this indirect method could suffer from systemic influences.

Altogether, it is not obvious which hypoxia biomarker will be the most useful for patient selection in clinical trials. All available methods and biomarkers to assess tumor hypoxia have their advantages and disadvantages. Since tumor hypoxia is heterogeneous and dynamic, probably a combination of biomarkers is necessary to identify patients with hypoxic tumors. Furthermore, biomarkers for specific prodrug-activating reductases, intrinsic sensitivity to the drug warhead and DNA repair status should also be taken into account in order to select a patient population that is expected to take full advantage from the HAP therapy.

5. Clinical trial design optimization

Biomarker assessment has not always been incorporated into the design of studies that attempt to modulate the hypoxic tumor fraction in some way, although in some cases retrospective information about hypoxia status is available. For example, a retrospective study on the DANish Head And Neck Cancer (DAHANCA) 5 trial using the aforementioned 15 hypoxia gene signature showed that the radiosensitizer nimorazole only benefitted the hypoxic group [107]. Recently, this signature has been technically validated (DAHANCA 30) [110] and was found suitable for use in prospective studies. Furthermore, the DAHANCE 33 image-guided dose-escalation radiotherapy trial (NCT02976051) aimed at improved patient stratification based on FAZA PET imaging has started. The multicentric EORTC-1219-ROG-HNC/DAHANCA 29 (NCT02661152) phase III trial in which non-responders (patients with less hypoxic tumors) are randomized based on the 15-gene signature for treatment with or without nimorazole is currently recruiting patients to verify whether the routinely used nimorazole can be omitted for less hypoxic HNSCC patients [111]. With regard to HAPs, plasma OPN levels were retrospectively detected in a subset of patients from a phase III trial of carboplatin/paclitaxel with or without TPZ [112]. It was shown that pretreatment plasma levels of OPN were significantly associated with patient response and that it thus may have utility as a prognostic biomarker. Additionally, Rischin and colleagues [113] applied FMISO-PET hypoxia scans pre- and mid-treatment in a subset of patients from a larger multi-centered phase II tirapazamine trial. The risk of loco-regional failure in the patients with hypoxic tumors was significantly higher when patients were treated by standard therapy, compared with combination therapy. This evidence shows the importance of identifying hypoxic tumors when treating patients with HAPs. However, the Phase III studies for TPZ and TH-302 did not include any up-front strategies for biomarker-guided identification of tumor hypoxia. Instead, all of these Phase III trials have roughly the same trial design, in which all patients were randomized to either receive standard of care, or the combination of standard treatment with the HAP. These retrospective results strongly indicate that assessing the hypoxia status of tumors is of utmost importance to guide the success of HAPs. Incorporation of hypoxia status assessment in future clinical trials will not only increase the
chance of demonstrating the beneficial effect of HAPs, but fewer patients will be needed to do so.

Regardless of the biomarker that will be used, a clear threshold must be established in order to categorize patients as biomarker-positive or biomarker-negative. This should be done in a Phase II setting since these studies are designed to determine whether new treatments show promising effects for further testing in Phase III trials. As such, a single arm study design is usually sufficient for Phase II studies. However, this becomes less straightforward in treatment regimens that are expected to have an effect in a selected (e.g., biomarker-positive) patient population only. This is also important for an adequate assessment of the expected treatment effect for a Phase III study and subsequently an adequate power and sample size determination. Biomarker-adaptive designs or biomarker-stratified Phase II studies have been developed to address the issue of identifying a possible biomarker-positive threshold and have been recently reviewed [114]. Some examples worth mentioning are the Multi-arm multi-stage, the Adaptive parallel Simon two-stage, and the Tandem two-stage design. In Multi-arm multi-stage designs [115], multiple treatment arms are tested simultaneously, but some are dropped early for futility. The different arms can be made up of different treatment regimens, but also of different patient groups with respect to biomarker classification. In the first stage of the Adaptive parallel Simon two-stage design [116], two parallel phase II studies are performed. In case of efficacy in both arms, biomarker selection will be stopped in the second stage. In case of efficacy in the biomarker-positive group only, the second stage will be completed with inclusion of biomarker-positive patients only. In the Tandem two-stage design [117], a phase II study is started with an unselected patient population. If the treatment appears effective after the first stage, the study is completed by including patients from the unselected population. If the first stage is unsuccessful, a second trial is started in a selected population.

The practical implications of these designs can be shown by an example. As a reference, a standard optimal two-stage phase II design is used [118]. Alpha and beta for all calculations are set at 0.05 and 0.20 respectively. Median survival of a poor (reference) treatment is determined at 12 months and the hazard ratio (HR) of a more active treatment is assumed to be 0.60. This corresponds to survival probabilities at 12 months of 0.50 for the reference treatment and 0.66 for a more active treatment. The sample size for the first stage of the study is 21 patients. If 11 or fewer patients respond to the treatment, the study is terminated. If the trial goes on to the second stage, a total of 72 patients will be studied. If the total number responding is less than or equal to 42, the new treatment is rejected. When applying this example on a Multi-arm multi-stage design, the sample size will depend on the number of arms. Each arm will have at most 72 patients, but it is likely that some arms will be terminated after 21 patients. In the adaptive parallel Simon two-stage design, two Phase II studies are started. After 21 patients per study for the first stage, one of the trials is stopped. The other is continued until 72 patients have been included. This brings the total number to 93 patients. In the Tandem two-stage design, the number of patients depends on the number of times that the first stage is unsuccessful. So at the very least 72 patients will be included, and an additional 21 for each first stage that is terminated. Thus establishing a threshold for biomarker positive patient selection can increase the number of patients needed in a Phase II study. However, adequate patient selection based on the predictive or prognostic evidence of a biomarker can increase power and/or decrease the required sample size for a subsequent Phase III trial.

Multiple reviews have addressed the numerous Phase III trial designs in which biomarkers can be incorporated [119–121]. For hypoxia-activated prodrugs we propose the targeted or enriched design, in which only the biomarker-positive patients are randomized between standard treatment and standard treatment in combination with the HAP (Fig. 1). This design is commonly used and is appropriate when there is preliminary evidence to suggest that treatment benefit is only expected in biomarker-positive patients. When an appropriate cut-off point has been established (in the aforementioned Phase II study), the study is very efficient, increases the power and, above all, requires a small sample size. By using biomarker-stratification, the HR is expected to be much lower than in a randomize-all design, which affects the number of patients needed for the study in order to demonstrate a significant difference of the treatment regimen. Fig. 2 shows numbers of patients needed for different HRs and different power levels assuming a median survival control of 12 months. Based on retrospective information about hypoxia status in previous studies [122,123], we presume that a decline of HR of 0.3 is feasible when HAP treatment is only given to hypoxia-positive patients instead of to the whole group of patients. When assuming a HR of 0.8 for the whole group of patients, this means that 109 patients would be needed for this phase III study compared to 951 when all patients are randomized (with a power of 0.9). This huge difference in total number of patients needed in Phase III more than compensates for the extra patients needed in the Phase II biomarker optimization phase. Even with lower differences in HR, still a large difference in patients needed to show a significant effect of the therapy exists. This implies a shorter time period to complete the study and assists patients to avoid potentially futile treatment regimens, an ethically responsible approach that will lead to reduced costs.

6. Conclusion

The high incidence of tumor hypoxia in cancer and its associated poor prognosis justifies expansion of ongoing efforts to
address this unmet need. Hypoxia is arguably the best validated target in oncology yet to be addressed by a successful therapy. Surprisingly, 50 years of HAP design has failed to yield a clinically approved agent, although only a handful of HAPs have actually advanced to the clinic during this period. The reasons for this failure are multifaceted, some of which are addressed in this review. Current phase III studies have omitted to stratify patients based on the hypoxia-status of the tumor. A biomarker-stratified enriched study design, with upfront assessment of the hypoxia biomarker threshold, will increase the chance to prove the beneficial effect of HAPs with fewer patients needed and subsequent implementation in clinical practice. The oncology community has the requisite tools to achieve success and they should be utilised constructively.

Declarations of interest

None.

Acknowledgements

This work was supported by Worldwide Cancer Research (15-0345) and EUROSTARS (COMPACT – E11993).

Appendix A: Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ctro.2019.01.005.

References


[47] Stokes AM, Hart CP, Quares JJ. Hypoxia imaging with pet correlates with


