(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 February 2024 (15.02.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/033599 A1

(51) International Patent Classification:
B01J 19/00 (2006.01)

(21) International Application Number:
PCT/GB2023/051595

(22) International Filing Date:
19 June 2023 (19.06.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2211750.1 11 August 2022 (11.08.2022) GB

(71) Applicant: **EVONETIX LTD** [GB/GB]; 9a Coldham's Business Park, Norman Way, Cambridge Cambridgeshire CB1 3LH (GB).

(72) Inventors: **SANCHES-KUIPER, Dr. Raquel Maria**; c/o Evonetix Ltd, 9a Coldham's Business Park, Norman Way, Cambridge Cambridgeshire CB1 3LH (GB). **HAYES, Dr. Matthew James**; c/o Evonetix Ltd, 9a Coldham's Business Park, Norman Way, Cambridge Cambridgeshire CB1 3LH (GB). **CROSBY, Dr. Stuart Richard**; c/o Evonetix Ltd, 9a Coldham's Business Park, Norman Way, Cambridge Cambridgeshire CB1 3LH (GB).

(74) Agent: **BERRYMAN, Robert**; D Young & Co LLP, 120 Holborn, London EC1N 2DY (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY,

(54) Title: REACTION DURATION CONTROL FOR REACTION PERFORMED ON BIOMOLECULE SAMPLES

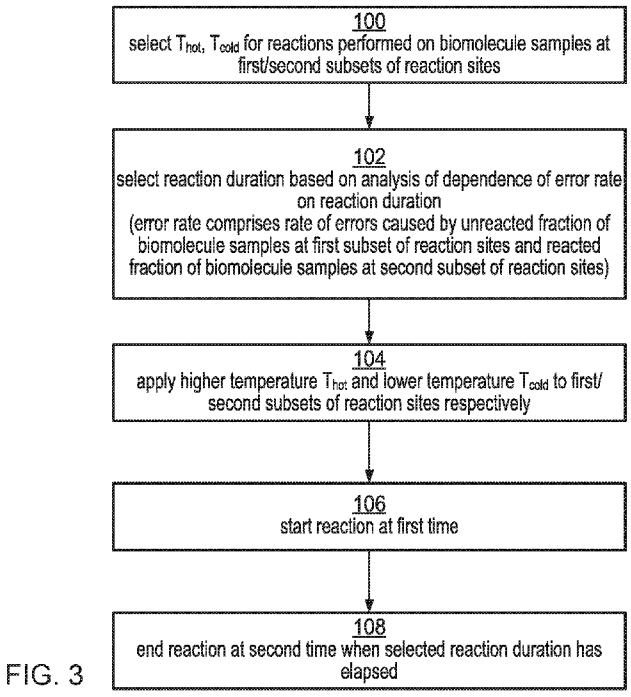

FIG. 3

(57) Abstract: A reaction duration is controlled for a reaction performed on biomolecule samples on an apparatus comprising temperature control circuitry to independently control temperature for respective reaction sites. A higher temperature is applied to a first subset of reaction sites and a lower temperature is applied to a second subset of reaction sites. The reaction is started at a first time and ended at a second time when the reaction duration has elapsed since the first time. The reaction duration is selected depending on an analysis of dependence of an error rate on the reaction duration, the error rate comprising a rate of errors caused by an unreacted fraction of biomolecule samples at the first subset of reaction sites that remain unreacted at the second time and a reacted fraction of biomolecule samples at the second subset of reaction sites that have already reacted at the second time.

WO 2024/033599 A1

[Continued on next page]

MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

REACTION DURATION CONTROL FOR REACTION PERFORMED ON BIOMOLECULE SAMPLES

The present technique relates to the field of biomolecule reactions.

Biomolecule samples, e.g. short or long strands of single-stranded or double-stranded DNA, RNA or other nucleic acid, or peptides, may be processed in a reaction. The reaction could be performed for a variety of purposes, such as synthesis, purification, or error correction of groups of samples of nucleic acid sequences. The yield of the reaction (ratio of useful output material (that meets a desired requirement for the reaction result) to input material) may depend on the rate of errors occurring in the reaction. Errors may occur for various reasons, causing some samples to not react correctly. It may be desirable to provide a technique for improving the yield of a reaction performed on biomolecule samples.

At least some examples provide a method of controlling a reaction duration for a reaction performed on biomolecule samples on an apparatus comprising a plurality of reaction sites each to support a set of biomolecule samples and temperature control circuitry to independently control temperature for respective reaction sites, the method comprising:

applying a higher temperature to a first subset of reaction sites and a lower temperature to a second subset of reaction sites;

starting the reaction at a first time; and

ending the reaction at a second time when the reaction duration has elapsed since the first time;

wherein the reaction duration is selected depending on an analysis of dependence of an error rate on the reaction duration, the error rate comprising a rate of errors caused by an unreacted fraction of biomolecule samples at the first subset of reaction sites that remain unreacted at the second time and a reacted fraction of biomolecule samples at the second subset of reaction sites that have already reacted at the second time.

At least some examples provide an apparatus comprising:

a plurality of reaction sites each to support a set of biomolecule samples;

temperature control circuitry to independently control temperature for respective reaction sites; and

control circuitry to control the apparatus to:

apply a higher temperature to a first subset of reaction sites and a lower temperature to a second subset of reaction sites;

start the reaction at a first time; and

end the reaction at a second time when a reaction duration has elapsed since the first time;

wherein the control circuitry is configured to select the reaction duration depending on an analysis of dependence of an error rate on the reaction duration, the error rate

comprising a rate of errors caused by an unreacted fraction of biomolecule samples at the first subset of reaction sites that remain unreacted at the second time and a reacted fraction of biomolecule samples at the second subset of reaction sites that have already reacted at the second time.

At least some examples provide a computer program comprising instructions which, when executed by an apparatus comprising at least one processor for executing the instructions, a plurality of reaction sites each to support a set of biomolecule samples and temperature control circuitry to independently control temperature for respective reaction sites, controls the apparatus to perform the method described above. The computer program may be stored on a storage medium, which may be a non-transitory or transitory storage medium.

Further aspects, features and advantages of the present technique will be apparent from the following description of examples, which is to be read in conjunction with the accompanying drawings, in which:

Figures 1 and 2 illustrate an example of an apparatus having reaction sites and temperature control circuitry;

Figure 3 illustrates a method of controlling reaction duration;

Figure 4 illustrates how the unreacted fraction of biomolecule samples at hotter and colder reaction sites varies with increasing reaction duration;

Figure 5 illustrates how an error rate caused by an unreacted fraction of samples at the hotter reaction site and a reacted fraction of samples at the colder reaction site varies with increasing reaction duration, indicating that there is a target reaction duration at which the error rate is minimised;

Figure 6 illustrates an example showing dependence of the target reaction duration on reaction conditions (e.g. pH);

Figure 7 illustrates dependence of the target reaction duration on samples with different bases; and

Figure 8 illustrates average fraction of reactions in error over a synthesis cycle for all bases.

For some reactions performed on biomolecule samples, it can be desirable to be able to select particular subsets of samples for undergoing the reaction while other subsets of samples are left unreacted. While it is possible to use pooling or other physical separation techniques to distinguish the subsets of samples which are to undergo a given stage of the reaction process, this can require either manual intervention, or more complex hardware, for transporting samples between pools between different stages of a reaction process, which can increase operating costs and the overall reaction time.

A temperature-based approach for controlling which subsets of biomolecule samples are processed in a given reaction step can be useful for addressing these issues. The biomolecule

samples may be disposed on an apparatus comprising a number of reaction sites which each can support a respective set of biomolecule samples. Temperature control circuitry can independently control temperature for respective reaction sites. This allows a first subset of reaction sites to be set to a higher temperature than a second subset of reaction sites. As reaction rate is dependent on temperature, the difference in temperature between sites can be used to make it more probable that the reaction is performed on the set of biomolecule samples at the first (hotter) subset of reaction sites than at the second (colder) subset of reaction sites. Hence, it is possible to perform targeted control of reactions applied to individual sets of samples without needing as much physical separation between the different sets of samples.

However, the inventors recognised that, for any given difference in temperatures between the first and second subsets of reaction sites, at a given time in the reaction there is an unreacted fraction of biomolecule samples at the first subset of reaction sites (that it is preferred to have reacted) and a reacted fraction of biomolecule samples at the second subset of reaction sites (that it is preferred not to have reacted). These fractions of samples represent a source of error causing a reduction in yield in the process of which the reaction forms a part.

The inventors recognised that the unreacted fraction of biomolecule samples at the first subset of reaction sites will decrease over time, while the reacted fraction of biomolecule samples at the second subset of reaction sites will increase over time, and that there is an intermediate time window at which the error rate (caused by the sum of the unreacted fraction of biomolecule samples at the first subset of reaction sites and the reacted fraction of biomolecule samples at the second subset of reaction sites) will be lower than windows of time earlier in the reaction and later in the reaction. Therefore, by controlling the reaction duration for which the reaction is allowed to run after being started, based on analysis of the dependence of this error rate on reaction duration, it is possible to select a reaction duration which reduces the error rate and hence increases yield.

Hence, a method comprises applying a higher temperature to a first subset of reaction sites and a lower temperature to a second subset of reaction sites; starting the reaction at a first time; and ending the reaction at a second time when the reaction duration has elapsed since the first time. The reaction duration is selected depending on an analysis of dependence of an error rate on the reaction duration, the error rate comprising a rate of errors caused by an unreacted fraction of biomolecule samples at the first subset of reaction sites that remain unreacted at the second time and a reacted fraction of biomolecule samples at the second subset of reaction sites that have already reacted at the second time. This helps to improve yield in the reaction. This approach is particularly useful for reactions applied to biomolecule samples because such processing can be extremely sensitive to loss of yield, e.g. due to incorporation or hybridisation errors. Hence, any improvement in yield by reducing a source of error can be valuable in increasing the amount of final product produced for a given amount of input material.

The reaction duration may be selected depending on a prediction of a target reaction duration at which the error rate is predicted to be minimised. The prediction of the target reaction duration may be based on the higher and lower temperatures applied to the first and second subsets of reaction sites, and on one or more reaction kinetic parameters which influence the dependence of reaction rate on temperature. By considering a prediction of the particular duration at which the error rate is predicted to be minimised when setting the reaction duration, the reaction can be halted at the second time, when it is predicted that the error rate is likely to be lower, hence improving yield.

The prediction of the target reaction duration could be obtained in different ways. For some prediction models, it is possible to derive a mathematical expression expressing the target reaction duration in dependence on reaction rate parameters dependent on reaction rates, at the higher temperature and the lower temperature respectively, of the reaction at the first subset of sites and the second subset of sites. As indicated in the Appendix below, this expression may be based on solving a differential equation derived from an expression for the error rate, to identify the target reaction direction at which the error rate is minimised.

For example, the target reaction duration t_{min} may be predicted according to an expression equivalent to:

$$t_{min} = \frac{\ln(k_{hot}/k_{cold})}{k_{hot} - k_{cold}}$$

where:

k_{cold} is the reaction rate constant for the reaction at the second subset of reaction sites at the lower temperature T_{cold}, and

k_{hot} is the reaction rate constant for the reaction at the first subset of reaction sites at the higher temperature T_{hot}. This expression is derived in the Appendix below. It will be appreciated that there are a number of different ways of defining alternative expressions which are completely equivalent to each other, so the particular expression used by a particular system may vary while nevertheless being equivalent to the expression defined above (for example, swapping k_{cold} with k_{hot} in the expression above gives exactly the same value for t_{min}).

In other examples, rather than using a mathematical expression for the reaction duration itself, the target reaction duration may be predicted based on numerical analysis of error rates calculated for different reaction durations using an error rate expression defining the error rate in dependence on the reaction duration, the higher temperature, the lower temperature and one or more reaction kinetic parameters influencing reaction rate for the reaction at the first subset of sites and the second subset of sites. The reaction kinetic parameters may themselves depend on a number of factors such as reaction conditions, the particular bases used for the samples, etc. Hence, depending on the sophistication of the model used to determine the reaction kinetic parameters, it may be that a mathematical solution to the error rate minimisation problem is hard to solve analytically, but can be solved numerically by calculating or plotting the error rate for

different times, and observing from the numeric results the time at which error rate is minimised. Hence, it is not essential to use an analytical expression to determine the prediction of the target reaction duration.

In another example, the analysis of dependence of an error rate on the reaction duration could be by conducting experiments to test the yield obtained in the reaction when the reaction is run for different reaction durations, and selecting the reaction duration that gives the highest yield. These experiments are based on analysis of error rate because it is recognised that there will be a time at which the yield will be highest given the rate of errors caused by the unreacted fraction of biomolecule samples at the first subset of reaction sites and the reacted fraction of biomolecule samples at the second subset of reaction sites. Hence, in some cases, it may not be necessary to calculate any analytical expression for the error rate, if the target reaction duration can be obtained experimentally by sweeping different reaction durations and comparing yields.

The analysis of the error rate's dependence on reaction duration could be performed on the fly at the time of performing reaction itself. However, this analysis could also already have been performed in the past, and at the time of the reaction, previously stored data may be read to determine the reaction duration to use for a certain reaction step. For example, a lookup table could be provided to look up the reaction duration to be applied given values for the higher temperature, lower temperature and an explicit or implicit identification of the reaction kinetic parameters (the reaction kinetic parameters may be implicitly represented based on a reaction ID). Alternatively, in some cases the reaction durations to be used for a particular set reactions may already have been fully calculated in advance (based on analysis of error rate as discussed above), so the lookup table may associate a reaction identifier identifying the particular reaction step to be performed with the reaction duration to use for that step. Hence, at the time of the reaction itself, control circuitry controlling the reaction may read out the reaction duration to be used for a particular reaction step from a previously stored data structure indicating the control settings for the reaction process, where that data structure is derived depending on analysis of the dependence of error rate on the reaction duration as described above.

Although error can be minimised by setting the reaction duration to be equal to the target reaction duration, it is not essential for the reaction duration to be set precisely equal to the target reaction duration. A window of time comprising a margin either side of the target reaction duration may be acceptable in order to reduce error rates sufficiently compared to reaction durations outside that window. In particular, it may be that there are a number of different reactions occurring in parallel, or the reaction may be applied to a number of different biomolecule samples with different sequences, and so as the overall reaction time may be controlled in common to those different reactions or different samples, it may be that choosing the optimum reaction duration for one reaction may cause the reaction duration to be a little off the minimum reaction duration for another reaction. Also, control tolerance limitations in the granularity with which the

reaction duration can be controlled may also mean that it may not be possible to set the reaction duration exactly equal to the target reaction duration. In another example, it may be that the operating costs associated with a slightly reduced yield associated with not operating exactly at the target reaction duration are offset by the cost savings achieved by having a shorter average reaction duration (hence, improving throughput and producing more product in a given time), so some implementations may choose to set the reaction duration slightly earlier than the target reaction duration so that error rate can be reduced to some extent but this is balanced with the improved efficiency of reactions being shorter on average.

Hence, in some examples the reaction duration is selected to be within ±25% of the target reaction duration. More specifically, the reaction duration could be selected to be within ±20% of the target reaction duration, or within ±15% of the target reaction duration, or within ±10% of the target reaction duration, or within ±5% of the target reaction duration, or within ±2 of the target reaction duration, or within ±1% of the target reaction duration.

In other examples, the reaction duration is selected to be, within bounds of control tolerance, the target reaction duration. This can allow a greater improvement in yield compared to examples which allow a wider margin around the target reaction duration.

The analysis of dependence of error rate on the reaction duration may account for a relative number of reaction sites in the first subset and the second subset. This reflects that if there are a greater number of reaction sites in one of the first subset and the second subset than the other, the relative contribution to error rate caused by the unreacted fraction of biomolecule samples at the first subset of reaction sites may be either greater than, or less than, the contribution to error rate caused by the reacted fraction of biomolecule samples at the second subset of reaction sites. Hence, by weighting the prediction of target reaction duration based on the relative numbers of reaction sites in the two subsets, this can provide a more precise prediction of the reaction duration at which overall error rate can be reduced.

The analysis of dependence of error rate on the reaction duration may also account for different reaction kinetic parameters for the first subset of reaction sites and the second subset of reaction sites, the reaction kinetic parameters comprising one or more parameters influencing dependence of reaction rate on temperature. For example, if the first subset of reaction sites have a different set of biomolecule samples compared to the second subset reaction sites, then the reactions on the samples may experience different reaction rates even when set to the same temperature. By modelling the reaction kinetic parameters more precisely to reflect the particular reactions occurring at each subset of sites, this can allow a more precise estimate of the reaction duration at which error rate can be reduced.

Similarly, there may be different groups of reaction sites within the first subset or different groups of reaction sites within the second subset that may experience different reaction kinetic parameters (e.g. because those groups of reaction sites host samples with different bases).

Hence, the analysis of dependence of error rate on the reaction duration may account for at least one of: different reaction kinetic parameters for respective groups of reaction sites within the first subset of reaction sites; and different reaction kinetic parameters for respective groups of reaction sites within the second subset of reaction sites.

As mentioned above, where different sites are associated with different reaction kinetic parameters, but reaction duration is controlled in common for multiple sites, it may not be possible to set the reaction duration to the optimum time for each site. Hence, the reaction duration may be selected depending on a prediction of a target reaction duration at which a combined error rate is predicted to be minimised, the combined error rate comprising a total of error rates for different reactions having different reaction kinetic parameters performed at the first subset of reaction sites and the second subset of reaction sites, the reaction kinetic parameters comprising one or more parameters influencing dependence of reaction rate on temperature. By considering the combined effect on error rate for a number of different reactions having different reaction kinetic parameters the overall error can be reduced.

For the above examples referring to different reaction kinetic parameters for different reactions, one reason for the different reaction kinetic parameters can be that the reactions are performed on different biomolecule samples with different bases (e.g. when the biomolecule samples are nucleic acid fragments or oligonucleotides). Hence, some implementations may consider the particular base sequences of samples involved in the reactions when performing the analysis of dependence of error rate on reaction duration.

It will be appreciated that different implementations may choose more precise or less precise models of the effects of particular bases on error rate, as a model which fully considers the precise base sequences of every sample involved in the reactions across the apparatus as a whole may be too complex to analyse and so a level of approximation may be used, for example approximating certain similar reactions as having the same reaction kinetic parameters even if the reaction kinetic parameters would actually vary slightly depending on factors not considered in the model.

The time at which the reaction is started and stopped may be controlled in different ways. For example, starting the reaction could be controlled by starting supply of one or more reagents for the reaction, and/or chemically removing a protecting group from the biomolecule samples. Ending the reaction may be controlled by halting supply of, or removing, the one or more reagents; and forming a protecting group on the biomolecule samples (including re-forming the protecting group, if the protecting group was previously present at an earlier stage of processing). It will be appreciated that, in practice, it may take a certain amount of time for reaction to start after the reaction starting control action is taken, and it similarly may take a certain amount of time for the reaction to end after the reaction ending control action is taken. Also, in practice not all the samples undergoing the reaction may start or end the reaction at the same time, so there may be

distribution in actual reaction durations experienced by different subsets of biomolecule samples. Nevertheless, the timing of the reaction starting control action and reaction ending control action may be selected so that the duration between them is controlled based on the analysis of dependence of error rate on reaction duration, so that the average reaction duration seen by a set of samples as a whole is more likely to correspond more closely to the target reaction duration at which the error rate is predicted to be minimised, helping to improve yield.

The biomolecule samples could be any samples comprising biomolecules. The biomolecules may be, for example, biopolymers (such as polynucleotides or polypeptides). For example, the biomolecules may comprise at least one of: double-stranded DNA, RNA or another nucleic acid; single-stranded DNA, RNA or another nucleic acid; synthesis initiators for DNA, RNA or another nucleic acid (e.g. the synthesis initiators could be pre-synthesised short biomolecules, chemical linkers and/or solid supports functionalized with reaction groups); and peptides, DNA peptide conjugates, or peptide hybrids (or conjugates) with another nucleic acid. The biomolecules could be oligonucleotides (which could be single-stranded or double-stranded nucleic acids).

The technique described above could be used for controlling reaction duration for any reaction applied to biomolecule samples. For example, the reaction may be part of a synthesis process for synthesizing the biomolecule samples. For example, the reaction may comprise a deprotection reaction to remove a protecting group from the biomolecule samples, a release reaction to release biomolecule samples from a reaction site surface, or a denaturation reaction to denature nucleic acid samples (e.g. to cause a double-stranded nucleic acid sample to separate into single-stranded nucleic acid samples). It may be desirable for such reactions to be applied preferentially to one set of samples and not to another. Temperature control can be a useful way of controlling the selective application of the reaction to a set of samples, and the reaction duration control method described above can be useful for reducing error rate and hence improving yield.

One use of nucleic acid samples can be as data storage. Information can be encoded in the sequence of base pairs that form a nucleic acid molecule, such as DNA, RNA or XNA. A desired sequence can be synthesized *de novo* in a factory or lab, to generate molecules which represent the data to be encoded. For example, with DNA molecules, each base in the sequence is one of four possible choices (adenine, cytosine, guanine or thymine), and so each base may encode two bits of information. Nucleic acid based data storage is an attractive option because the physical density is very high. It is possible to use a temperature-controlled mechanism for addressing a memory implemented using nucleic acid samples stored on reaction sites of an apparatus as discussed above. For example, PCT patent publication WO 2020/021221 A1 describes a method for temperature-controlled addressing of data stored using nucleic acid samples. The reactions involved in the addressing of the data storage may have a reaction rate

varying based on temperature. When it is desired to select one set of samples for reacting while leaving another set of samples unreacted, temperature differences between the sites used to store these samples can be used to select which set of samples is accessed. Hence, in some cases, the biomolecule samples comprise nucleic acid samples and the reaction forms part of storage or retrieval of data stored in nucleic-acid-based data storage. For example, the temperature-dependent reactions described in WO 2020/021221 A1 (the contents of which are hereby incorporated by reference) can have their reaction duration controlled based on the technique discussed above.

Figures 1 and 2 illustrate an apparatus 2 on which oligonucleotide processing reactions can be performed. As shown in Figure 1, the apparatus has a fluid flow path 4 across the top of the device. A fluid flow element (e.g. a pump) is provided to control the flow of fluid through the fluid flow path 4. A number of reaction sites (active thermal sites) 6 are provided at various locations across the plane of the temperature control device 2. The top of each reaction site 6 may include a reaction surface (e.g. a gold cap) on which oligonucleotide samples, such as single-stranded or double-stranded nucleic acid fragments, can be synthesised, attached, cleaved, or processed. Each reaction site 6 corresponds to part of a level surface, so that there is no physical barrier between adjacent reaction sites 6. Each reaction site 6 has a heating element 7 (temperature control circuitry, e.g. a resistive heater) provided below the reaction site surface to apply heat to the corresponding part of the fluid flowing over that site, to control the temperature of the fluid. The heat applied at each site is variable, based on control signals provided by an external controller. Hence, the temperature control circuitry is able to independently set the temperature for each reaction site 6. As shown in Figure 2, the reaction sites 6 are arranged in a two-dimensional matrix (grid), arranged in two or more rows (lanes) 9 where the lane/row direction is parallel to the direction that fluid flows through the fluid flow path 4. The regions lying between the active thermal sites 6 form one or more passive thermal regions 8 which do not comprise any heating element, but provide passive cooling by conducting heat away from the fluid towards the substrate 10 of the device 2. The length x of each active thermal site 6 in the row direction is longer than the length y of each passive thermal region 8 lying between a pair of adjacent active thermal sites 6 in the same row. The thermal resistance of the material provided below each active thermal site 6 in a direction perpendicular to the substrate may be greater than the thermal resistance in the direction perpendicular to the substrate of the material provided below each passive thermal region 8, which is helpful for improving the temperature range supported for a given maximum power provided by each heating element 7. As shown in Figure 1, a cooling mechanism 12 may be provided to cool the substrate 10 to act as a heat sink. Further information about the apparatus shown in Figures 1 and 2, including an example of a process for manufacturing it and a control model for controlling the heating elements 7, can be found in WO 2018/104698 A1 (the contents of which are hereby incorporated by reference).

Hence, the apparatus 2 is a thermal array allowing individual control of temperature at different regions across the two-dimensional surface of a substrate.

With such an apparatus 2, the reagents for a reaction can be supplied in the fluid travelling along the fluid flow path, to react with biomolecule samples disposed on the respective reaction sites. By setting the temperatures of respective sites, the reaction can be controlled to take place at a faster rate at one site compared to another, to provide selective control of which sites undergo the reaction. For example, the reaction could be any of:

- a reaction to remove a protective group from biomolecule samples at selected sites 6, e.g. to allow for the subsequent addition of bases to extend a nucleic acid sample to form a longer chain of bases. The protective group may be desired to be removed only from samples at certain sites, to avoid samples in other sites being extended with subsequently added bases.
- a reaction to release biomolecule samples from a reaction site surface, so that they can be transported to a different site in the flowing fluid or removed from the reaction apparatus 2 altogether. It may be that only selected sites 6 are desired to be subject to release of the biomolecule samples, and others should remain bound to their respective sites.
- a denaturation reaction to denature (separate into single strands) double-stranded nucleic acid samples, which could be useful for error detection or purification of a set of samples for removing samples which have incorporation or hybridisation errors. It may be that the denaturation operation is desired to be applied only to selected sites 6.
- a reaction for retrieving or storing data stored in nucleic-acid-based data storage, where the sequence of bases of nucleic acid samples disposed at the respective sites represents data values which can be read or written to by performing addressing reactions such as those described in WO 2020/021221 A1 mentioned above. In that document, a read operation comprises setting a target site to be read to a higher temperature than other sites not to be read, to provide a greater probability of the double-stranded nucleic acid molecules separating into single-stranded nucleic acid molecules at the target site than at the other sites, allowing the separated molecules to be amplified and sequenced to read the encoded data represented by those molecules. A write operation comprises setting a target site to be written at a lower temperature than other sites, to provide a greater probability of new double-stranded nucleic acid molecules forming at the target site compared to other sites (i.e. the probability of molecules undergoing a separation reaction is higher at the hotter other sites). Hence, the reaction involved in the read/write operation could be the reaction whose reaction duration is controlled based on the examples set out in this application.

For all these examples, the reaction duration control discussed in the following examples could be applied.

Figure 3 illustrates a method of controlling reaction duration for a reaction performed on biomolecule samples disposed on an apparatus 2 having multiple reaction sites 6 with independent control of the temperature of each site, such as the example shown in Figures 1 and 2. The biomolecule samples could be single-stranded or double-stranded sequences of DNA, RNA, XNA or other nucleic acid, for example, including nucleic acids with modified (non-standard) bases, or could be synthesis initiators for DNA, RNA or another nucleic acid, or samples of peptides, DNA peptide conjugates, or peptide hybrids with RNA or another nucleic acid.

At step 100, higher and lower temperatures T_{hot} and T_{cold} are selected for reactions performed on biomolecule samples at first and second subsets of reaction sites respectively. For example, the temperatures T_{hot} and T_{cold} could be selected depending on the purpose of the reaction being performed, and in some cases based on the specific bases included in the samples at the respective sites. Different reactions, and different stages of reaction within a process such as a nucleic acid synthesis process, may involve different settings for the higher and lower temperatures.

At step 102, a reaction duration is selected based on an analysis of dependence of error rate on reaction duration. The error rate comprises a rate of errors caused by an unreacted fraction of biomolecule samples at the first subset of reaction sites and a reacted fraction of biomolecule samples at the second subset of reaction sites. Various examples of analyzing the dependence of error rate on reaction duration and selecting the reaction duration are set out below. In general, the reaction duration is selected with an aim to reduce (e.g. minimise) the error rate.

Steps 100 and 102 could be done on the fly at the time when the subsequent steps of performing the reaction are carried out, or could have been carried out in advance so that at the time of the reaction itself, the reaction is controlled based on previously calculated parameters defining the higher and lower temperatures T_{hot} and T_{cold} and the reaction duration to use for a particular stage of the reaction.

At step 104, the higher and lower temperatures T_{hot} and T_{cold} are applied to the first and second subsets of reaction sites respectively. For example, control circuitry (not shown in Figure 1) may adjust control signals to the heaters 7 under the reaction sites to cause the temperature at sites 6 to be at the higher and lower temperatures T_{hot} and T_{cold} for the first and second subsets of reaction sites respectively.

At step 106, the reaction is started at a first time. For example, the reaction could be started by starting the flow of reagents in the fluid supplied via the fluid flow path 4 for a lane of reaction sites, or by chemically removing a protecting group from biomolecule samples (where that protecting group is a group with a structure which inhibits the reaction from taking place).

The elapse of time during the reaction is monitored, and at step 108, at a second time at which the reaction duration selected at step 102 is determined to have been elapsed, the reaction

is ended. For example, the reaction could be ended by halting the flow of reagents or allowing the protecting group to form on the biomolecule samples.

The accuracy of control by heat depends on the temperature dependency of the rate of reaction between hot and cold sites. Many textbooks give a rule of thumb that states that most reaction rates in the solution-phase double for every 10°C temperature increase, but there is in fact a range. The rate of reaction, k, (or equivalently the half-life, which is -ln(0.5)/k) is given by the Arrhenius equation:

$$k = Ae^{\frac{-E_a}{RT}}$$

where R = 8.314 (the gas constant), T is the absolute temperature (in units of Kelvin), E_a is activation energy (in units of J/mol) and A is a pre-exponential factor (in units of s^{-1}). The activation energy and pre-exponential factor are dependent on properties of the specific reaction being performed (e.g. the type of reaction, the base sequences in the samples being reacted, the way the reaction is controlled, etc.).

This implies that:

$$\frac{k(T+10)}{k(T)} \approx e^{\left(\frac{10E_a}{RT^2}\right)}, \quad \frac{t_{half}(T+10)}{t_{half}(T)} \approx e^{\left(\frac{-10E_a}{RT^2}\right)}$$

The relative reaction rates between two temperatures, and therefore the controllability by heat, depends on the activation energy, which should be maximized (energies around 120kJ/mol to 130kJ/mol are practical, which corresponds to a reaction rate increase of 4 to 5.5 times per 10°C). This finite increase over practical temperature ranges (e.g. 20°C to 90°C in some examples) leads to errors due to the proportion of the reaction that was supposed to occur at a hot site but did not occur plus the proportion of reaction that was not supposed to occur at a cold site but did occur. As the reaction proceeds exponentially to completion, this error rate depends not only on the reaction kinetics, but also on how long the reaction is allowed to proceed, *t*.

This is shown in Figure 4. The fraction of unreacted samples remaining at a given time can be expressed as a function of the half-life (-ln(0.5)/k), where k depends on temperature according to the Arrhenius equation shown above. The half-life (time taken for half the remaining samples to react at a given site) is shorter for hotter sites than for colder sites. Hence, as shown in Figure 4, the number of remaining samples at the hotter site (represented by line 152) decays faster than the number of remaining samples at the colder site (represented by line 150). Although line 150 appears to be linear in Figure 4, this is just due to the limited time window shown in the graph – if viewed over a long enough time period line 150 would show a similar exponential decay to line 152.

At a given time t (assuming the reaction starts at time 0), the remaining fraction of unreacted samples at the hotter site and colder site respectively can be expressed as:

remaining fraction of unreacted samples at hotter site: $\frac{1}{2^{\frac{t}{t_{hhot}}}}$

remaining fraction of unreacted samples at colder site: $\frac{1}{2^{\frac{t}{t_{hcold}}}}$

t_{hhot} = half-life at hot site = $\frac{-\ln(0.5)}{k_{hot}}$

t_{hcold} = half-life at cold site= $\frac{-\ln(0.5)}{k_{cold}}$.

The error rate caused by the relative reaction rates between the two temperatures can therefore be expressed as:

$$error\ rate = 1 - \frac{1}{2^{\frac{t}{t_{hcold}}}} + \frac{1}{2^{\frac{t}{t_{hhot}}}}$$

where $1 - \frac{1}{2^{\frac{t}{t_{hcold}}}}$ represents the contribution to error caused by the fraction of reacted samples which have already reacted at the colder site but should not have reacted (corresponding to the region above line 150 in the graph shown in Figure 4), and $\frac{1}{2^{\frac{t}{t_{hhot}}}}$ represents the contribution to the error rate caused by the fraction of unreacted samples which have not reacted at the hotter site but which should have reacted in an ideal reaction (corresponding to the region below line 152 in the graph shown in Figure 4). The region between lines 150 and 152 represents the useful products of the reaction – the fraction of reacted samples at the hotter site plus the fraction of unreacted samples at the colder site.

This error rate is dominated at short times by incomplete reaction of the hot sites and at long time by unwanted reaction at the cold sites. As shown in Figure 5, which plots the error rate against time from the start of the reaction, there is therefore a target reaction time that minimizes this error rate, which can be calculated either analytically or numerically, depending on the degree of complexity of the prediction model used to represent the kinetic parameters of the reactions taking place under shared reaction duration control, or can be determined by experiment, sweeping different reaction times for different instances of performing the reaction and determining which reaction time gives the highest yield.

The Appendix to this patent specification shows a derivation of a formula for calculating the target reaction duration at which the error rate is minimised, if it is approximated (for simplicity) that all hotter sites share the same parameters Ea and A for all reactions taking place at the hotter (first subset) sites and all colder sites share the same parameters Ea and A for all reactions taking place at the colder (second subset) sites, and that the volume of biomolecule samples undergoing the reaction (e.g. the number of reaction sites in each subset) are equivalent for both subsets. With this derivation, the target reaction duration t_{min} can be expressed as:

$$t_{min} = \frac{\ln\left(\frac{k_{hot}}{k_{cold}}\right)}{k_{hot} - k_{cold}}$$

and the minimum error rate achieved if the reaction is ended at that target reaction duration can be expressed as:

$$error\ rate|_{t=t_{min}} = 1 - \frac{1}{\frac{\alpha\, r}{2^{(r-1)}}} + \frac{1}{\frac{\alpha}{2^{(r-1)}}}$$

where $r = \frac{k_{cold}}{k_{hot}}$ and $\alpha = \frac{\ln r}{\ln 2}$.

It will be appreciated that the reaction rates k_{hot}, k_{cold} depend on temperature and the reaction kinetic parameters A, E_a as shown in the Arrhenius equation above.

As shown in the Appendix, an equivalent expression for t_{min} can also be obtained by swapping k_{hot} and k_{cold} – it is an arbitrary choice which expression is used. It will be appreciated that other equivalent expressions could also be used, e.g. by representing the ratio $\frac{k_{hot}}{k_{cold}}$ as r, or by representing the reaction duration in terms of the half-lives at the hot and cold sites rather than the reaction rates.

Hence, the target reaction duration can be predicted according to such an expression or an equivalent one. The reaction duration for a particular reaction may however not necessarily be set exactly to the target reaction duration t_{min} at step 102 of Figure 3, e.g. due to limitations in control tolerance in the control of reaction duration, or because other factors are considered (e.g. it may be preferable to reduce processing times and so a reaction duration slightly shorter than the target reaction duration may be chosen). However, it can be useful if the reaction duration is set relatively close to the target reaction duration, e.g. within 25% or closer as explained above. With this approach, given likely values for the reaction kinetic parameters, error rates as low as 0.1% per reaction are practical, with preferred reaction times around 5 to 10 minutes.

It is well understood that the activation energy of reactions can be manipulated in many different ways, such as by changing the structure of the reacting molecules, the solvent, reagents, catalyst or pH. Although the error rate is minimized by control of the activation energy, it can also be desirable to reduce the target reaction duration time to achieve the greatest speed, improving throughput, reducing operating costs. It is also highly desirable to distinguish between the reaction that requires accelerating, and any others in the process that can be accelerated by increased temperature, thus reducing the detrimental effect of side-reactions. This can be achieved by controlling the design of the reaction and the reagents involved so as to manipulate the Arrhenius pre-exponential factor, *A* whilst maintaining the high Ea required for low error. This pre-exponential factor depends on both the number of collisions between reacting molecules, and on whether the molecules are properly oriented when they collide, so by designing the reaction environment or the chemistry of the molecules involved appropriately to promote a greater number of correctly oriented collisions, this can help to adjust the preferred reaction duration towards shorter durations.

Nevertheless, for any given values of A and Ea, it is possible to find a reaction duration at which error can be reduced as discussed above, so it is not essential to further optimise the pre-

exponential factor and activation energy. It will be appreciated that specific values for A and Ea will depend on the particular reaction being performed. The optimisation of A and Ea is a separate problem from the problem of how to reduce errors for a given set of values for A and Ea, and the solution of controlling the reaction duration is an orthogonal solution to any other techniques for optimising A and Ea by controlling the design of the reaction and reagents.

The reaction kinetic parameters also depend on the environment in which the reaction is performed. For example, Figure 6 shows how the dependence of error rate on reaction time varies with the pH at which the reaction is performed. At different pH, the estimated values for A and E_a change, causing differences in the time at which error rate is predicted to be minimised and the minimum error rate achieved at that time.

As shown in Figure 7, the reaction kinetic parameters also depend on the bases involved in the reaction. For example, Figure 7 considers a deprotection reaction to remove a protecting group from a fragment of nucleic acid. It is modelled that the reaction kinetic parameters for the deprotection reaction depend on the base at the end closest to the protecting group (e.g. A, T, G or C for a DNA fragment), and that the reaction kinetic parameters A and E_a are independent of the bases in the remainder of the sequence. Hence, there are four possible bases involved in any given deprotection reaction. These are associated with different reaction kinetic parameters as shown in the legend of Figure 7, and this causes the target reaction time (at which error rate is minimised) and the minimum error rate at that target reaction time to vary for all four bases. In a group of reaction sites 6 sharing the same control of start/end times for the reaction, it is preferred to reduce the total error rate across all the reaction sites.

Hence, if it is assumed that, at both the hot and cold subsets of sites, the probability of the base on the end of the sequence nearest the protecting group being A, T, G or C is equal for all four bases, and assuming that any given reaction step is attempting to remove the protecting group only from selected reaction sites where the base on the end of the sequence nearest the protecting group is a specific one of A, T, G, or C, then there is a ¼ probability that any given site will be one of the hot sites and a ¾ probability that any given site will be one of the cold sites. We can therefore weight the error rate calculation to assume that there will be 3 times as many cold sites as hot sites on average, and calculate the average contribution to error rate at both the hot and cold sites, averaged across the different values for the reaction kinetics for the respective bases (ε represents the error rate):

$$\varepsilon_{overall} = 3\overline{\varepsilon_{cold}} + \varepsilon_{hot}$$

$$\varepsilon_{overall} = \frac{3}{4}\sum_{i=[ATGC]}\left(1 - \frac{1}{2^{\frac{t}{t_{hcold,i}}}}\right) + \frac{1}{4}\sum_{i=[ATGC]}\left(\frac{1}{2^{\frac{t}{t_{hhot,i}}}}\right)$$

where:

$t_{hhot,i}$ = half-life at hot site where the end base is base i (selected from A, T, G, C)

$t_{hcold,i}$ = half-life at cold site where the end base is base i (selected from A, T, G, C).

Of course, if it was estimated that there would not be an equal probability of each base occurring as the end base, then the expression above could be weighted based on the relative probabilities of each base occurring as the end base.

For such a more complex expression for the overall error rate, it becomes difficult to solve the differential equation for determining the target reaction duration t_{min} analytically, but it can be done numerically by calculating values for $\varepsilon_{overall}$ for different values of t using the expression above, and then checking which time gives the minimum value for $\varepsilon_{overall}$.

Figure 8 shows an example of plotting, against reaction time, the average fraction in error over a synthesis cycle for all bases, based on the expression for $\varepsilon_{overall}$ shown above and example values for the reaction kinetic parameters. Hence, a reaction time can be chosen based on the time at which the overall error rate resulting from different base sequences can be predicted to be minimised.

For other reactions, the kinetic parameters for the reaction might depend on a longer sequence of bases than a single base, and so specific patterns of base sequences might be considered. Alternatively, different patterns could be approximated with an average set of reaction parameters. It will be appreciated that the model used could be more or less complex.

Hence, in summary, by analysing the way the contribution to error caused by the difference in reaction rate between hotter and colder sites depends on the time for which the reaction is allowed to run, a reaction duration can be selected which allows the error rate to be reduced, improving yield of the reaction.

Although illustrative embodiments of the invention have been described in detail herein with reference to the accompanying drawings, it is to be understood that the invention is not limited to those precise embodiments, and that various changes and modifications can be effected therein by one skilled in the art without departing from the scope of the invention as defined by the appended claims.

Appendix – derivation of expression for target reaction duration t_{min}

$$t_{hhot} = \text{half-life at hot site} = -\frac{\ln\left(\frac{1}{2}\right)}{k_{hot}} = \frac{\ln 2}{k_{hot}}$$

$$t_{hcold} = \text{half-life at cold site} = -\frac{\ln\left(\frac{1}{2}\right)}{k_{cold}} = \frac{\ln 2}{k_{cold}}$$

$$error\ rate = 1 - \frac{1}{2^{\frac{t}{t_{hcold}}}} + \frac{1}{2^{\frac{t}{t_{hhot}}}} = 1 - \left(\frac{1}{2}\right)^{\frac{t}{t_{hcold}}} + \left(\frac{1}{2}\right)^{\frac{t}{t_{hhot}}}$$

$$\frac{d(error\ rate)}{dt} = -\frac{1}{t_{hcold}}\left(\frac{1}{2}\right)^{\frac{t}{t_{hcold}}}\ln\left(\frac{1}{2}\right) + \frac{1}{t_{hhot}}\left(\frac{1}{2}\right)^{\frac{t}{t_{hhot}}}\ln\left(\frac{1}{2}\right)$$

$$\left.\frac{d(error\ rate)}{dt}\right|_{t=t_{min}} = 0$$

$$\left.\frac{d(error\ rate)}{dt}\right|_{t=t_{min}} = -\frac{1}{t_{hcold}}\left(\frac{1}{2}\right)^{\frac{t_{min}}{t_{hcold}}}\ln\left(\frac{1}{2}\right) + \frac{1}{t_{hhot}}\left(\frac{1}{2}\right)^{\frac{t_{min}}{t_{hhot}}}\ln\left(\frac{1}{2}\right) = 0$$

$$\frac{1}{t_{hhot}}\left(\frac{1}{2}\right)^{\frac{t_{min}}{t_{hhot}}}\ln\left(\frac{1}{2}\right) = \frac{1}{t_{hcold}}\left(\frac{1}{2}\right)^{\frac{t_{min}}{t_{hcold}}}\ln\left(\frac{1}{2}\right)$$

$$\frac{1}{t_{hhot}}\frac{1}{2^{\frac{t_{min}}{t_{hhot}}}} = \frac{1}{t_{hcold}}\frac{1}{2^{\frac{t_{min}}{t_{hcold}}}}$$

$$\frac{2^{\frac{t_{min}}{t_{hcold}}}}{2^{\frac{t_{min}}{t_{hhot}}}} = \frac{t_{hhot}}{t_{hcold}}$$

$$2^{t_{min}\left(\frac{1}{t_{hcold}} - \frac{1}{t_{hhot}}\right)} = \frac{t_{hhot}}{t_{hcold}}$$

$$t_{min}\left(\frac{1}{t_{hcold}} - \frac{1}{t_{hhot}}\right) = log_2\left(\frac{t_{hhot}}{t_{hcold}}\right) = \frac{\ln\left(\frac{t_{hhot}}{t_{hcold}}\right)}{\ln 2}$$

$$t_{min} = \frac{\ln\left(\frac{t_{hhot}}{t_{hcold}}\right)}{\ln 2\left(\frac{1}{t_{hcold}} - \frac{1}{t_{hhot}}\right)} = \frac{\ln\left(\frac{t_{hhot}}{t_{hcold}}\right)t_{hhot}}{\ln 2\left(\frac{t_{hhot}}{t_{hcold}} - 1\right)}$$

$$\frac{t_{hhot}}{t_{hcold}} = \frac{\frac{\ln 2}{k_{hot}}}{\frac{\ln 2}{k_{cold}}} = \frac{k_{cold}}{k_{hot}}$$

$$t_{min} = \frac{\ln\left(\frac{k_{cold}}{k_{hot}}\right)t_{hhot}}{\ln 2\left(\frac{k_{cold}}{k_{hot}} - 1\right)} = \frac{\ln\left(\frac{k_{cold}}{k_{hot}}\right)\ln 2}{\ln 2\ k_{hot}\left(\frac{k_{cold}}{k_{hot}} - 1\right)} = \frac{\ln\left(\frac{k_{cold}}{k_{hot}}\right)}{k_{cold} - k_{hot}}$$

Note that, since ln(A/B) = -ln(B/A), the expression for tmin can be represented equivalently as either:

$$t_{min} = \frac{\ln\left(\frac{k_{cold}}{k_{hot}}\right)}{k_{cold} - k_{hot}}$$

or

$$t_{min} = \frac{\ln\left(\frac{k_{hot}}{k_{cold}}\right)}{k_{hot} - k_{cold}}$$

(swapping k_{hot} and k_{cold} changes the sign of the expressions in both the numerator and denominator, so the resulting value for t_{min} remains the same).

From the above we had:

$$t_{min} = \frac{\ln\left(\frac{t_{hhot}}{t_{hcold}}\right) t_{hhot}}{\ln 2 \left(\frac{t_{hhot}}{t_{hcold}} - 1\right)}$$

We can use this to calculate the error rate at the target reaction duration t_{min}:

We can simplify by defining $r = \frac{t_{hhot}}{t_{hcold}} = \frac{k_{cold}}{k_{hot}}$ and $\alpha = \frac{\ln r}{\ln 2}$, to get:

$$t_{min} = \frac{\alpha\, t_{hhot}}{(r-1)}$$

Hence:

$$error\ rate|_{t=t_{min}} = 1 - \frac{1}{2^{\frac{t_{min}}{t_{hcold}}}} + \frac{1}{2^{\frac{t_{min}}{t_{hhot}}}} = 1 - \frac{1}{2^{\frac{\alpha\, t_{hhot}}{t_{hcold}(r-1)}}} + \frac{1}{2^{\frac{\alpha}{(r-1)}}}$$

$$error\ rate|_{t=t_{min}} = 1 - \frac{1}{2^{\frac{\alpha\, r}{(r-1)}}} + \frac{1}{2^{\frac{\alpha}{(r-1)}}}$$

If r had been defined the other way round as $r = \frac{t_{hcold}}{t_{hhot}}$ then the minimum error rate would become:

$$error\ rate|_{t=t_{min}} = 1 - \frac{1}{2^{\frac{\alpha}{(r-1)}}} + \frac{1}{2^{\frac{r\alpha}{(r-1)}}}$$

CLAIMS

1. A method of controlling a reaction duration for a reaction performed on biomolecule samples on an apparatus comprising a plurality of reaction sites each to support a set of biomolecule samples and temperature control circuitry to independently control temperature for respective reaction sites, the method comprising:

applying a higher temperature to a first subset of reaction sites and a lower temperature to a second subset of reaction sites;

starting the reaction at a first time; and

ending the reaction at a second time when the reaction duration has elapsed since the first time;

wherein the reaction duration is selected depending on an analysis of dependence of an error rate on the reaction duration, the error rate comprising a rate of errors caused by an unreacted fraction of biomolecule samples at the first subset of reaction sites that remain unreacted at the second time and a reacted fraction of biomolecule samples at the second subset of reaction sites that have already reacted at the second time.

2. The method of claim 1, in which the reaction duration is selected depending on a prediction of a target reaction duration at which an error rate is predicted to be minimised.

3. The method of claim 2, in which the target reaction duration is predicted based on an expression expressing the target reaction duration in dependence on reaction rate parameters dependent on reaction rates, at the higher temperature and the lower temperature respectively, of the reaction at the first subset of sites and the second subset of sites.

4. The method of claim 3, in which the target reaction duration t_{min} is predicted according to an expression equivalent to:

$$t_{min} = \frac{\ln(k_{hot}/k_{cold})}{k_{hot} - k_{cold}}$$

where:

k_{cold} is the reaction rate constant for the reaction at the second subset of reaction sites at the lower temperature T_{cold}, and

k_{hot} is the reaction rate constant for the reaction at the first subset of reaction sites at the higher temperature T_{hot}.

5. The method according to claim 2, in which the target reaction duration is predicted based on numerical analysis of error rates calculated for different reaction durations using an error rate

expression defining the error rate in dependence on the reaction duration, the higher temperature, the lower temperature and one or more reaction kinetic parameters influencing reaction rate for the reaction at the first subset of sites and the second subset of sites.

6. The method of any of claims 2 to 5, in which the reaction duration is selected to be within ±25% of the target reaction duration.

7. The method of any of claims 2 to 6, in which the reaction duration is selected to be, within bounds of control tolerance, the target reaction duration.

8. The method of any preceding claim, in which the reaction duration is selected to be a reaction duration at which the error rate less than 5%.

9. The method according to any preceding claim, in which the analysis of dependence of error rate on the reaction duration accounts for a relative number of reaction sites in the first subset and the second subset.

10. The method according to any preceding claim, in which the analysis of dependence of error rate on the reaction duration accounts for different reaction kinetic parameters for the first subset of reaction sites and the second subset of reaction sites, the reaction kinetic parameters comprising one or more parameters influencing dependence of reaction rate on temperature.

11. The method according to any preceding claim, in which the analysis of dependence of error rate on the reaction duration accounts for at least one of:

different reaction kinetic parameters for respective groups of reaction sites within the first subset of reaction sites; and

different reaction kinetic parameters for respective groups of reaction sites within the second subset of reaction sites;

the reaction kinetic parameters comprising one or more parameters influencing dependence of reaction rate on temperature.

12. The method according to any preceding claim, in which the reaction duration is selected depending on a prediction of a target reaction duration at which a combined error rate is predicted to be minimised, the combined error rate comprising a total of error rates for different reactions having different reaction kinetic parameters performed at the first subset of reaction sites and the second subset of reaction sites, the reaction kinetic parameters comprising one or more parameters influencing dependence of reaction rate on temperature.

13. The method according to any of claims 10 to 12, in which the reactions having the different reaction kinetic parameters comprise reactions performed on different biomolecule samples with different bases.

14. The method according to any preceding claim, in which starting the reaction comprises at least one of:

starting supply of one or more reagents for the reaction, and

chemically removing a protecting group from the biomolecule samples.

15. The method according to any preceding claim, in which ending the reaction comprises at least one of:

halting supply of, or removing, the one or more reagents; and

forming a protecting group on the biomolecule samples.

16. The method according to any preceding claim, in which the biomolecule samples comprise at least one of:

double-stranded DNA, RNA or another nucleic acid;

single-stranded DNA, RNA or another nucleic acid;

synthesis initiators for DNA, RNA or another nucleic acid; and

peptides, DNA peptide conjugates, or peptide hybrids with another nucleic acid.

17. The method according to any preceding claim, in which the reaction is part of a synthesis process for synthesizing the biomolecule samples.

18. The method according to any preceding claim, in which the reaction comprises a deprotection reaction to remove a protecting group from the biomolecule samples.

19. The method according to any of claims 1 to 17, in which the reaction comprises a release reaction to release biomolecule samples from a reaction site surface.

20. The method according to any of claims 1 to 17, in which the reaction comprises a denaturation reaction to denature nucleic acid samples.

21. The method according to any of claims 1 to 16 in which the reaction forms part of storage or retrieval of data stored in nucleic-acid-based data storage.

22. An apparatus comprising:

a plurality of reaction sites each to support a set of biomolecule samples;

temperature control circuitry to independently control temperature for respective reaction sites; and

control circuitry to control the apparatus to:

apply a higher temperature to a first subset of reaction sites and a lower temperature to a second subset of reaction sites;

start the reaction at a first time; and

end the reaction at a second time when a reaction duration has elapsed since the first time;

wherein the control circuitry is configured to select the reaction duration depending on an analysis of dependence of an error rate on the reaction duration, the error rate comprising a rate of errors caused by an unreacted fraction of biomolecule samples at the first subset of reaction sites that remain unreacted at the second time and a reacted fraction of biomolecule samples at the second subset of reaction sites that have already reacted at the second time.

23. A computer program comprising instructions which, when executed by an apparatus comprising at least one processor for executing the instructions, a plurality of reaction sites each to support a set of biomolecule samples and temperature control circuitry to independently control temperature for respective reaction sites, controls the apparatus to perform the method of any of claims 1 to 21.

24. A storage medium storing the computer program of claim 23.

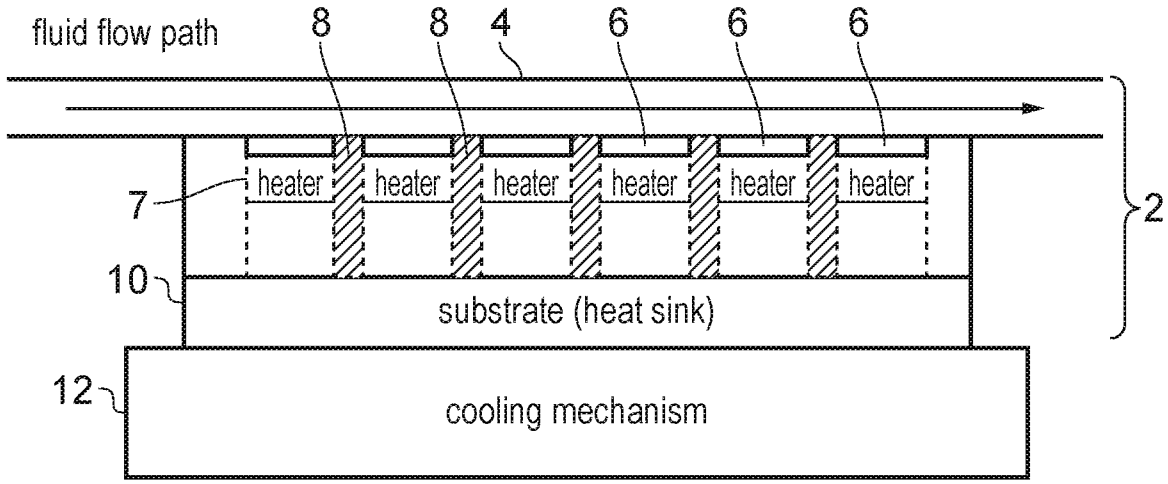
fluid flow path
8
8
4
6
6
6
heater
heater
heater
heater
heater
heater
7
2
10
substrate (heat sink)
12
cooling mechanism

FIG. 1

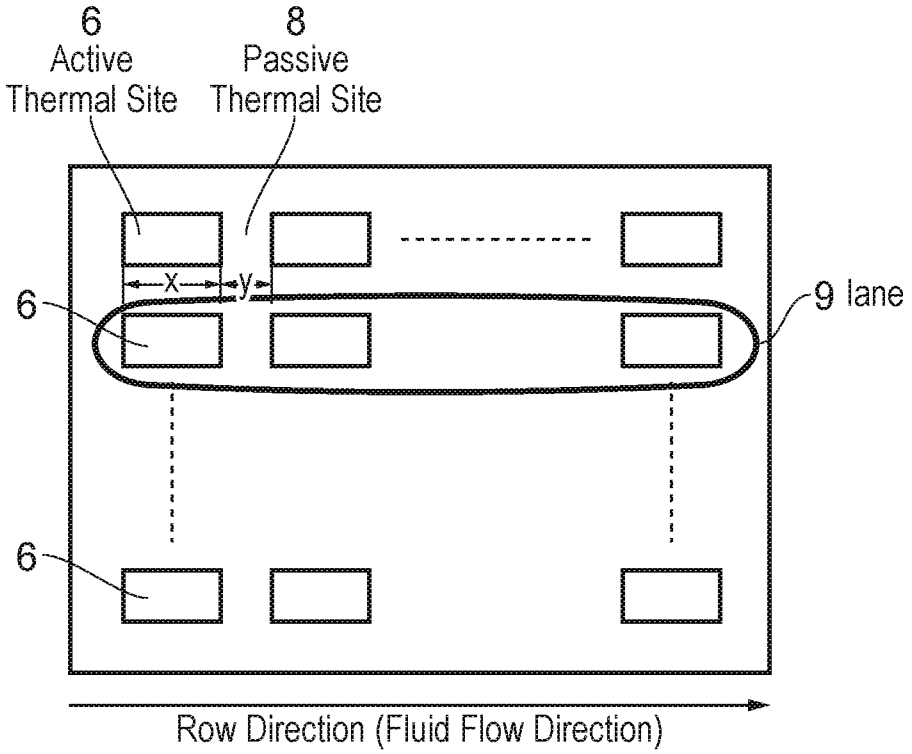
6
Active
Thermal Site
8
Passive
Thermal Site
x
y
6
9 lane
6
Row Direction (Fluid Flow Direction)

FIG. 2

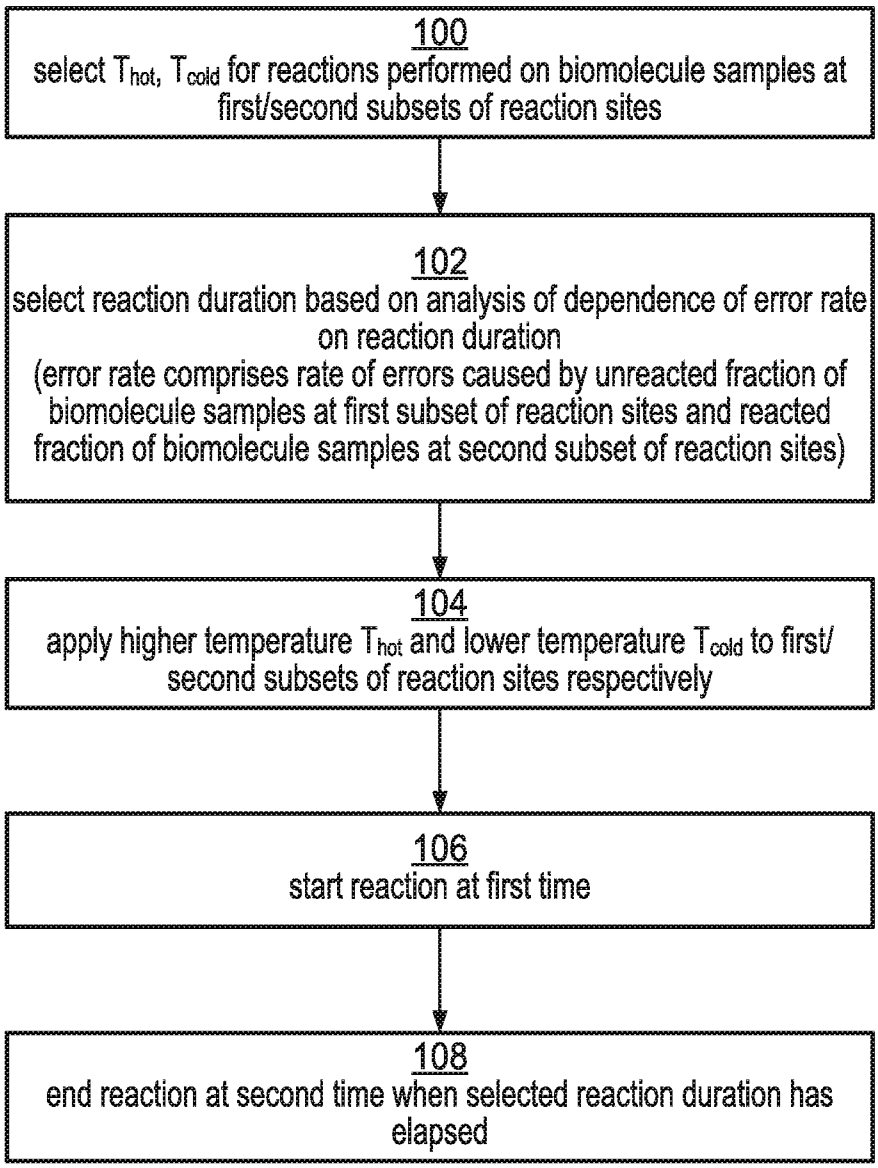
100
select T_{hot}, T_{cold} for reactions performed on biomolecule samples at first/second subsets of reaction sites
102
select reaction duration based on analysis of dependence of error rate on reaction duration
(error rate comprises rate of errors caused by unreacted fraction of biomolecule samples at first subset of reaction sites and reacted fraction of biomolecule samples at second subset of reaction sites)
104
apply higher temperature T_{hot} and lower temperature T_{cold} to first/second subsets of reaction sites respectively
106
start reaction at first time
108
end reaction at second time when selected reaction duration has elapsed

FIG. 3

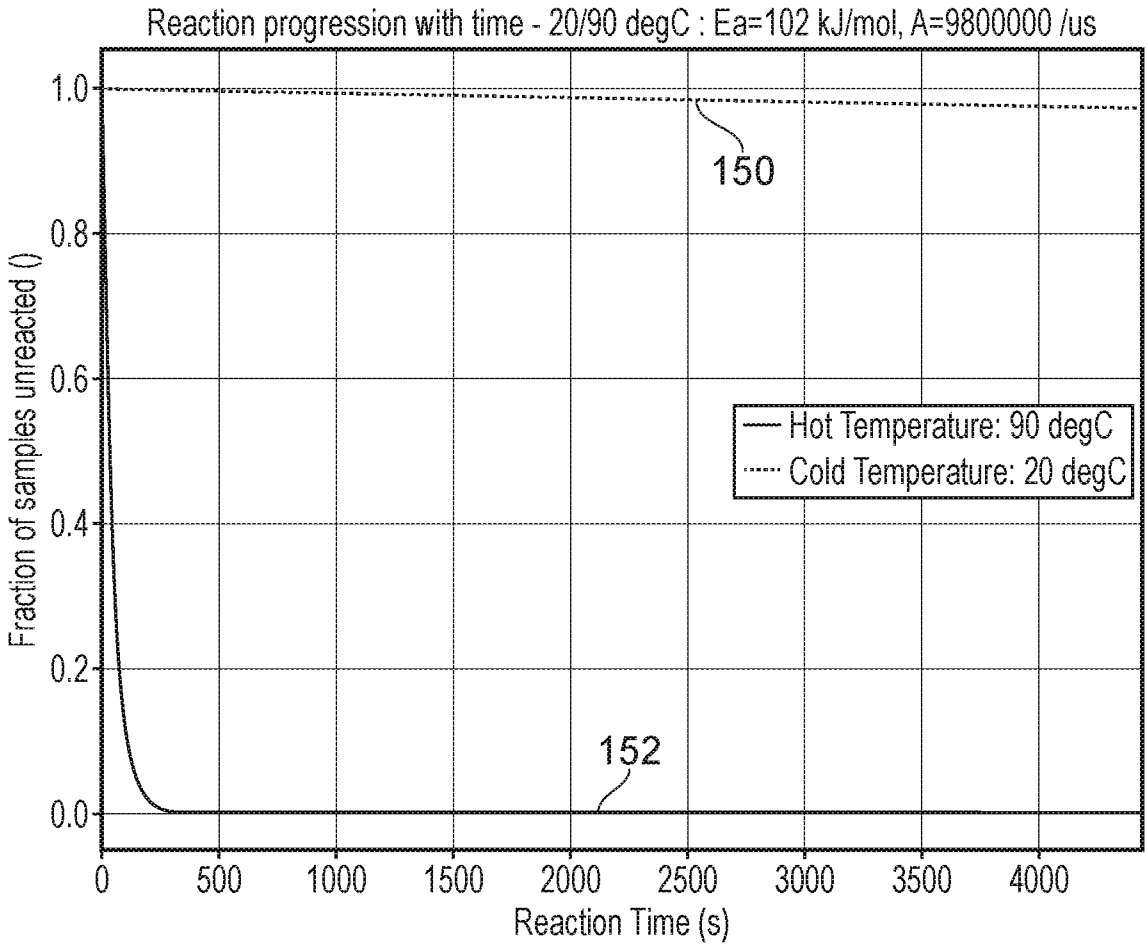
Reaction progression with time - 20/90 degC : Ea=102 kJ/mol, A=9800000 /us
Fraction of samples unreacted ()
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
2500
3000
3500
4000
Reaction Time (s)
150
152
Hot Temperature: 90 degC
Cold Temperature: 20 degC

FIG. 4

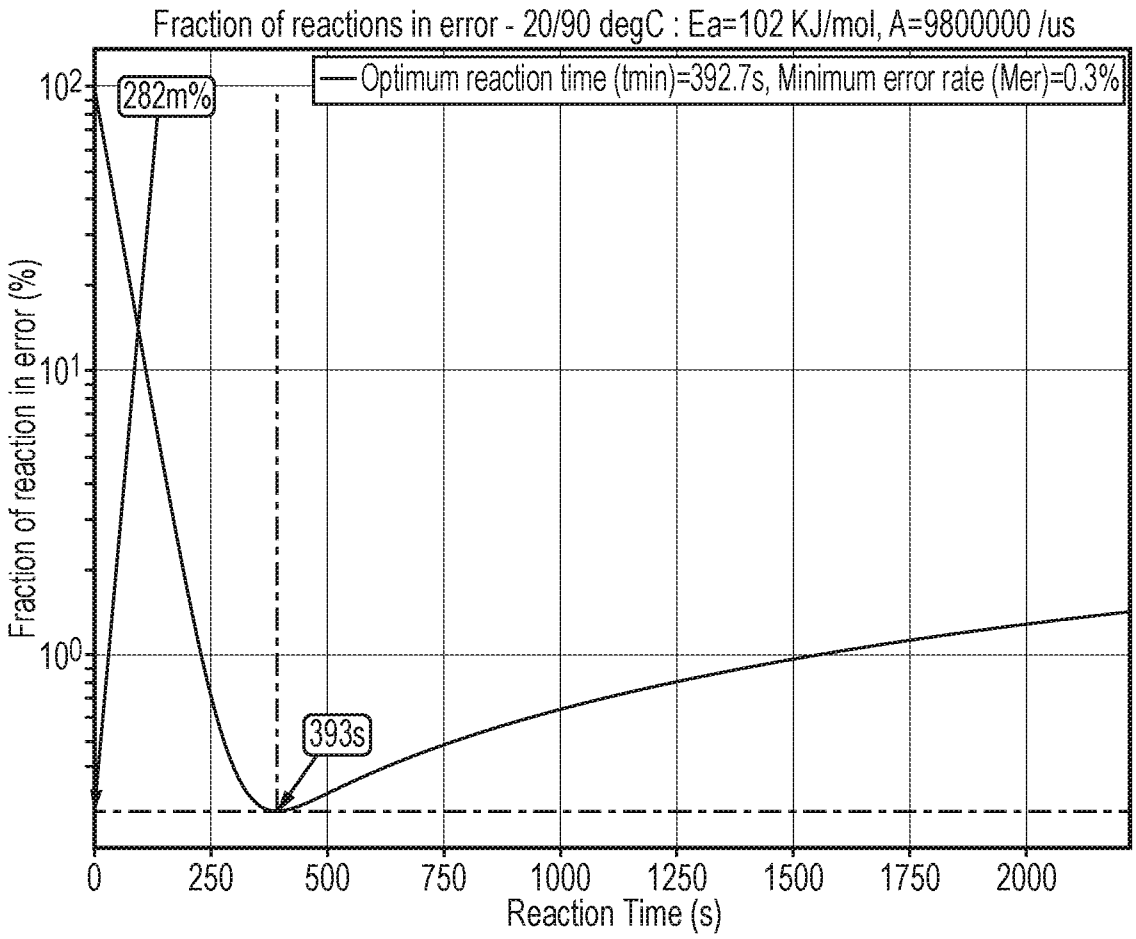
Fraction of reactions in error - 20/90 degC : Ea=102 KJ/mol, A=9800000 /us
Optimum reaction time (tmin)=392.7s, Minimum error rate (Mer)=0.3%
282m%
393s
Fraction of reaction in error (%)
10^2
10^1
10^0
0
250
500
750
1000
1250
1500
1750
2000
Reaction Time (s)

FIG. 5

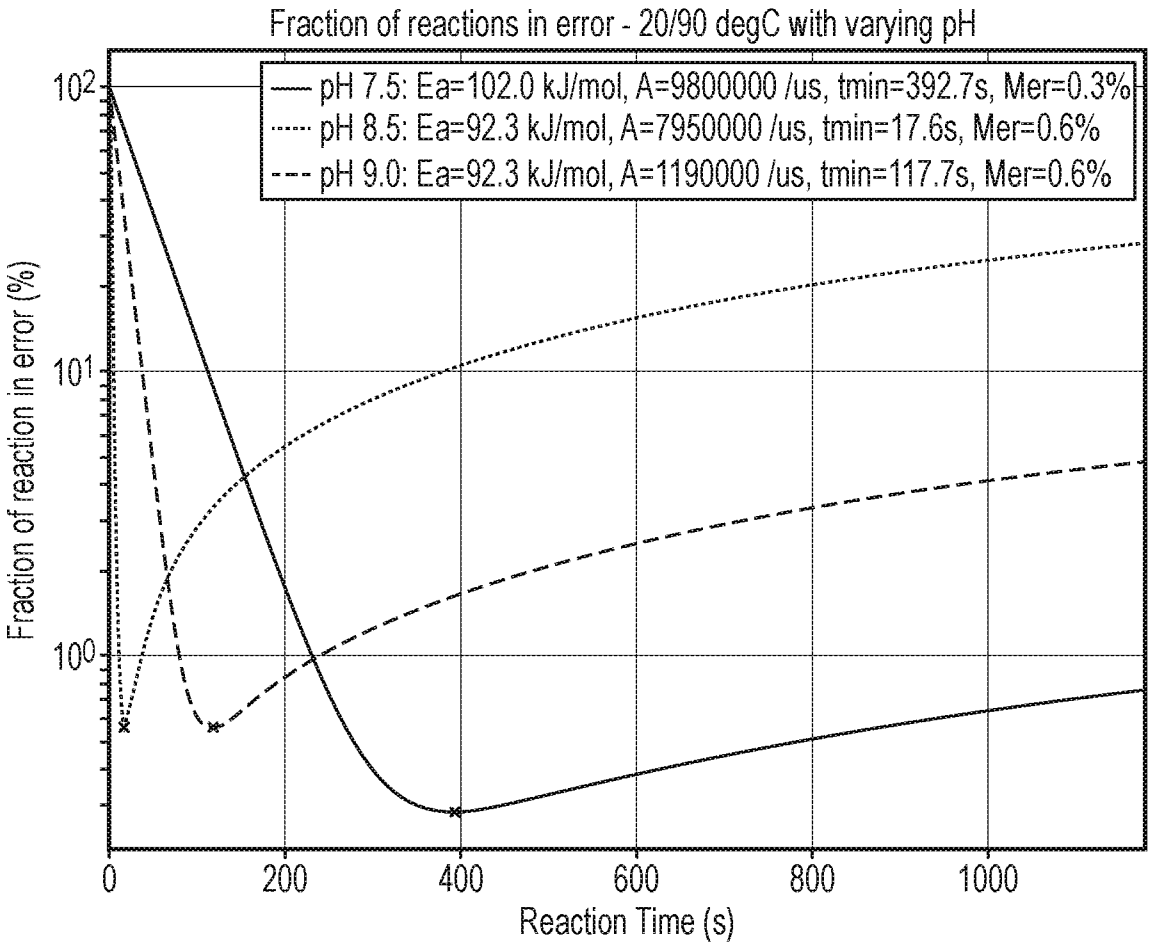
Fraction of reactions in error - 20/90 degC with varying pH
pH 7.5: Ea=102.0 kJ/mol, A=9800000 /us, tmin=392.7s, Mer=0.3%
pH 8.5: Ea=92.3 kJ/mol, A=7950000 /us, tmin=17.6s, Mer=0.6%
pH 9.0: Ea=92.3 kJ/mol, A=1190000 /us, tmin=117.7s, Mer=0.6%
Fraction of reaction in error (%)
10^2
10^1
10^0
0
200
400
600
800
1000
Reaction Time (s)

FIG. 6

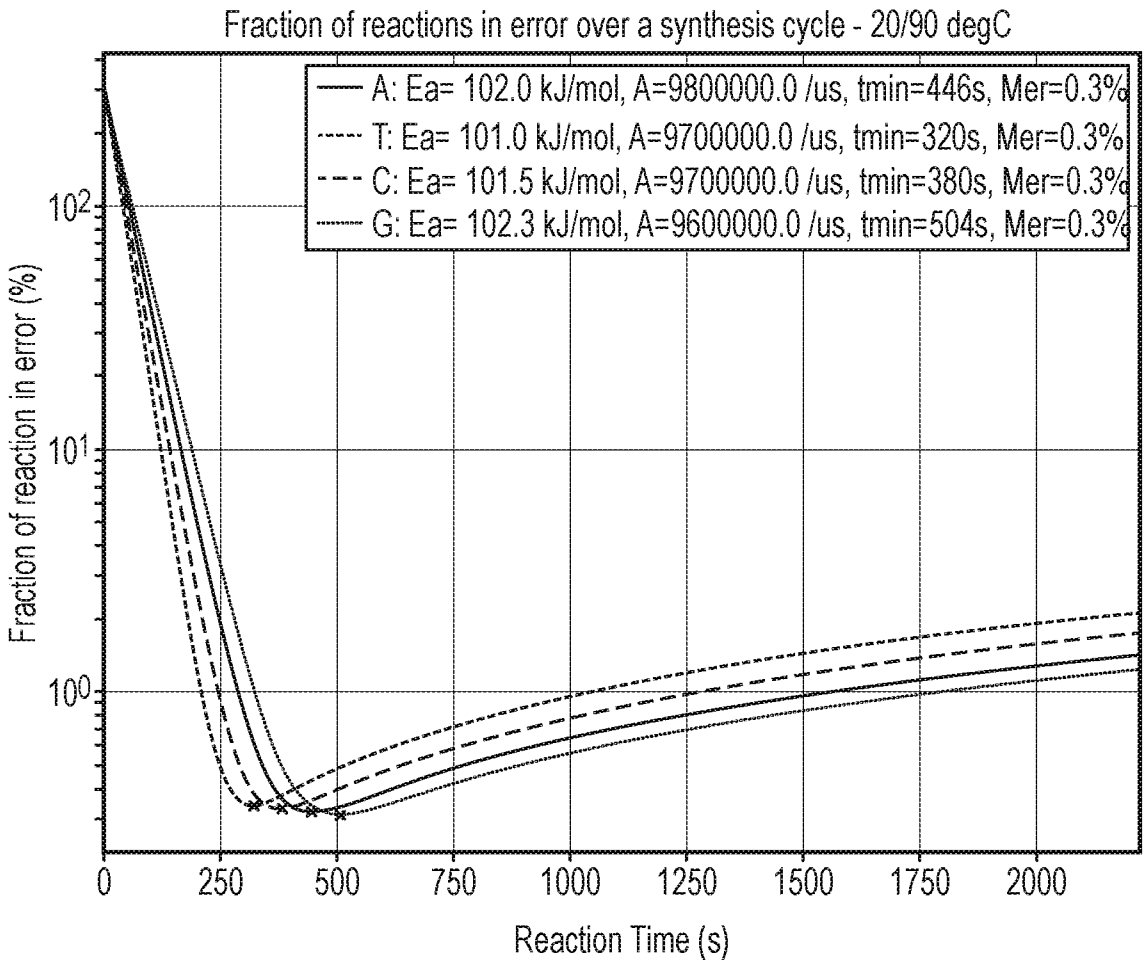
Fraction of reactions in error over a synthesis cycle - 20/90 degC
A: Ea= 102.0 kJ/mol, A=9800000.0 /us, tmin=446s, Mer=0.3%
T: Ea= 101.0 kJ/mol, A=9700000.0 /us, tmin=320s, Mer=0.3%
C: Ea= 101.5 kJ/mol, A=9700000.0 /us, tmin=380s, Mer=0.3%
G: Ea= 102.3 kJ/mol, A=9600000.0 /us, tmin=504s, Mer=0.3%
Fraction of reaction in error (%)
10^2
10^1
10^0
0
250
500
750
1000
1250
1500
1750
2000
Reaction Time (s)

FIG. 7

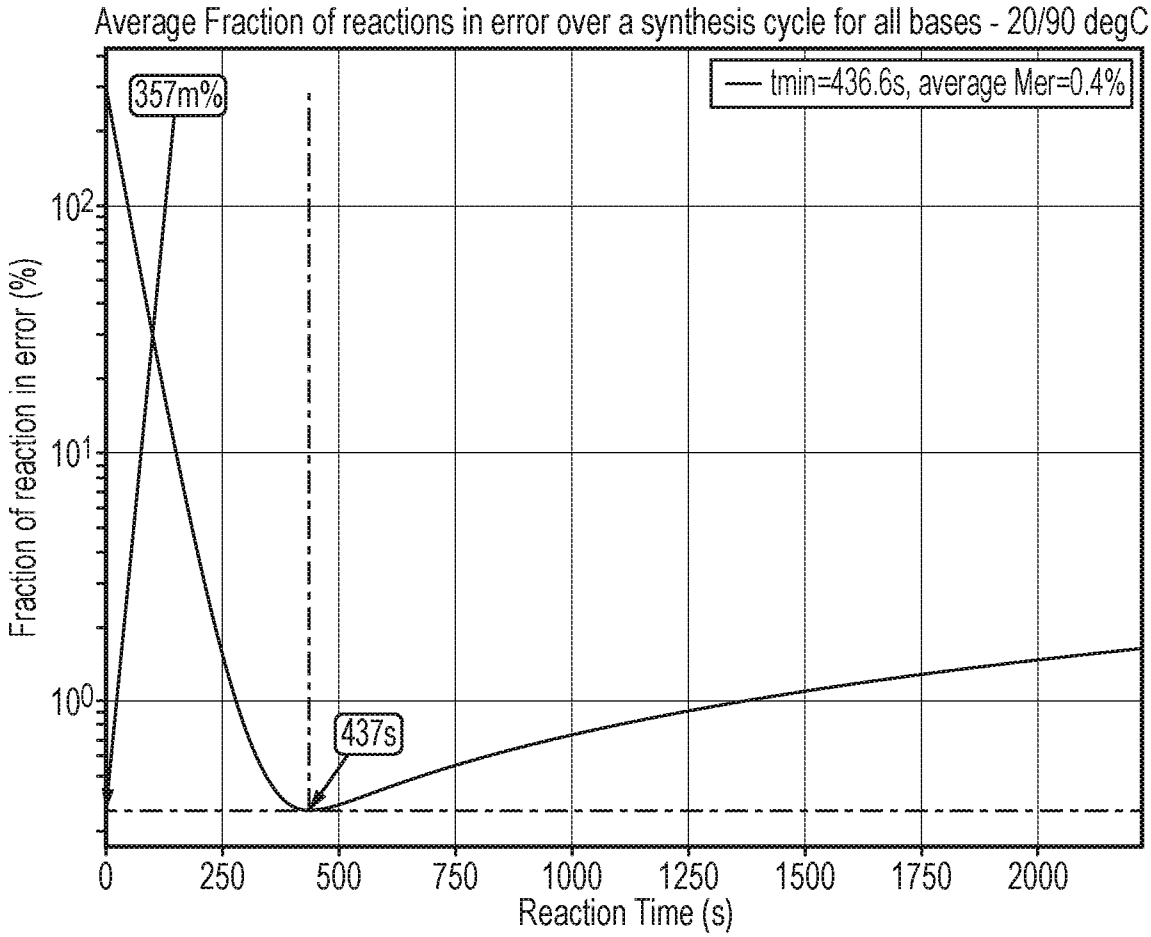
Average Fraction of reactions in error over a synthesis cycle for all bases - 20/90 degC
tmin=436.6s, average Mer=0.4%
357m%
437s
Fraction of reaction in error (%)
10^2
10^1
10^0
0
250
500
750
1000
1250
1500
1750
2000
Reaction Time (s)

FIG. 8

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2023/051595

A. CLASSIFICATION OF SUBJECT MATTER
INV. B01J19/00
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
B01J

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2020/021221 A1 (EVONETIX LTD [GB]) 30 January 2020 (2020-01-30) cited in the application abstract page 14, line 37 - page 15, line 34; figures 3, 4	1-24
A	US 2020/248254 A1 (HAYES MATTHEW JAMES [GB] ET AL) 6 August 2020 (2020-08-06) abstract paragraph [0118]; claim 1; figures 10, 11	1-24
A	WO 2018/104698 A1 (EVONETIX LTD [GB]) 14 June 2018 (2018-06-14) abstract page 12, line 6 - line 28; figures 1, 2 claims 1, 20	1-24

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
31 August 2023	26/09/2023

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Thomasson, Philippe

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2023/051595

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2020021221 A1	30-01-2020	CA 3105954 A1	30-01-2020
		EP 3827435 A1	02-06-2021
		GB 2576304 A	19-02-2020
		SG 11202100073U A	25-02-2021
		TW 202008381 A	16-02-2020
		US 2021280274 A1	09-09-2021
		WO 2020021221 A1	30-01-2020
US 2020248254 A1	06-08-2020	AU 2018343116 A1	26-03-2020
		CA 3077106 A1	04-04-2019
		CN 111164220 A	15-05-2020
		DK 3688189 T3	11-07-2022
		EP 3688189 A1	05-08-2020
		ES 2920355 T3	03-08-2022
		GB 2566986 A	03-04-2019
		GB 2568974 A	05-06-2019
		IL 273298 A	30-04-2020
		JP 7201674 B2	10-01-2023
		JP 2020537511 A	24-12-2020
		KR 20200058461 A	27-05-2020
		SG 11202002188Y A	29-04-2020
		US 2020248254 A1	06-08-2020
		WO 2019064006 A1	04-04-2019
WO 2018104698 A1	14-06-2018	AU 2017371801 A1	06-06-2019
		CA 3042610 A1	14-06-2018
		CN 110022981 A	16-07-2019
		EP 3551331 A1	16-10-2019
		GB 2557592 A	27-06-2018
		HU E057490 T2	28-05-2022
		IL 266815 A	31-07-2019
		JP 7022130 B2	17-02-2022
		JP 2020501271 A	16-01-2020
		KR 20190093595 A	09-08-2019
		TW 201828422 A	01-08-2018
		US 2019275527 A1	12-09-2019
		WO 2018104698 A1	14-06-2018