



(51) International Patent Classification:

G01N 31/00 (2006.01) G05B 15/02 (2006.01)  
G05B 13/02 (2006.01) G06F 19/00 (2011.01)

(21) International Application Number:

PCT/GB2017/051326

(22) International Filing Date:

12 May 2017 (12.05.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

1608683.7 17 May 2016 (17.05.2016) GB

(71) Applicant: TAP BIOSYSTEMS (PHC) LIMITED [GB/GB]; York Way, Royston Hertfordshire SG8 5WY (GB).

(72) Inventors: WALES, Richard; 18 Mill Hill Road, St Neots, Cambridgeshire PE19 7AQ (GB). COURTNEY, Patrick; Flat 2, 11 Mount Avenue, Ealing W5 1QB (GB).

(74) Agent: GILL JENNINGS & EVERY LLP; The Broadgate Tower, 20 Primrose Street, London EC2A 2ES (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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

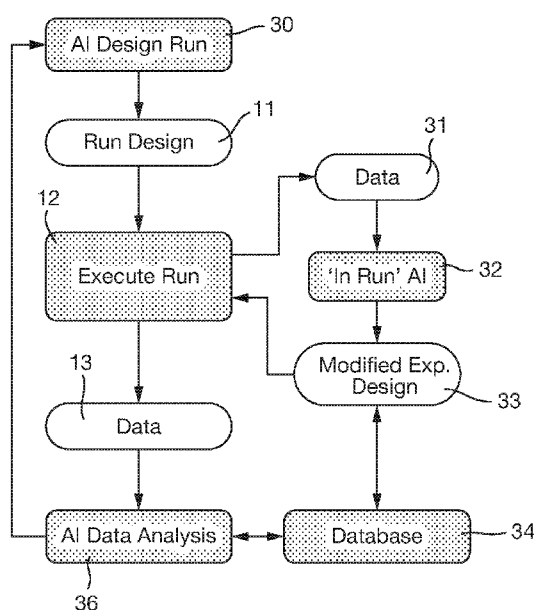
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, 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))

(54) Title: AUTOMATED BIOPROCESS DEVELOPMENT

Fig. 3



(57) Abstract: A method for automating process development in a bioprocessing environment is provided. The method comprising: executing a first experiment run according to a set of parameters; retrieving a first real-time set of data of the experiment run while the experiment run is being executed; retrieving a second real-time set of data of an experiment run being executed in parallel, analysing the retrieved first real-time set of data and the second real-time set of data to determine an adjusted set of parameters; and, modifying, based on the analysis, the parameters upon which the experiment run is being executed during execution of the run such that the run continues to be executed according to the modified set of parameters. A system, computer program and computer readable medium are also provided.

WO 2017/199006 A1

## AUTOMATED BIOPROCESS DEVELOPMENT

### Background

Bioprocess development typically involves statically designed experiments  
5 executed as a series of parallel runs in similar systems, in sequence on a single  
system or individually or in sequence in a parallel system capable of running  
parallel experiments but less than the total number of experiments required. The  
key material to be assessed in these experiments is valuable and limited in  
supply. Prior to experiment execution it is unknown how the material, for  
10 example a particular novel new candidate therapeutic protein or group of related  
such molecules, will respond to certain conditions. Indeed some materials can  
be particularly sensitive and very small changes in the material, for example a  
single amino acid in a protein can have a significant and unpredicted effect on  
the behaviour of the protein.

15

The current trends in development strategy add to the scale of the challenges in  
bioprocess development relating to the amount of material available for testing at  
early stages. One example of a group of such trends that create a specific  
challenge on material supply are:

- 20 i) an interest in understanding very early in the development process which of a  
number of identified therapeutic protein candidates are most suitable for the  
small set of manufacturing processes that a developer has adopted as their  
platform, often generally termed assessment of 'manufacturability';  
ii) a focus on injectable formats for biological drugs creating a demand for very  
25 high protein concentrations in order to achieve the required dose within a volume  
that can be tolerated by the patient for injection; and,  
iii) the increasing productisation of drugs in a competitive environment that adds  
to the constraints directing formulations to formats of high usability for patients.

30 At early stages of, or prior to process development, a critical point at which  
candidate molecule selection is carried out, typically all unit operations are  
inefficient and sub-optimal. One example of the type of issue that arises is  
illustrated by the assessment of the robustness of a protein drug candidate to  
manufacturing concentration and formulation processes involving ultra-filtration

and diafiltration (UF/DF), typically one of the last stages of manufacturing prior to filling of the delivery container. Available bench scale systems for UF/DF have minimum recirculating (operating) volumes in the region of 20-30 ml. The focus for many protein therapeutics is on injectable formulations where in order to keep the injected volume tolerable for the patient but deliver the required dose very high protein concentrations are required, e.g. 150-200mg/ml being targets for many developer companies. To achieve 200mg/ml in 25ml fluid requires 5g of protein. Whilst 2-3g/L of product can routinely be achieved as a titre in the manufacturing bioreactor for an optimised process, prior to or at the early stages of process development before optimisation yields are much lower, not only in the bioreactor but also in multiple subsequent separation and purification unit operations that precede the UF/DF step. As such a single trial might require an initial supply bioreactor run of 20L or greater. The cost of such scale runs is high, added to this the cost of subsequent work to separate the product from the cell culture and purify the protein (UF/DF is typically the final or penultimate step) is high, which makes significant demand on human and facility resources and of long duration. Where multiple different trial UF/DF runs are required to assess the performance of the candidate in the UF/DF process the scale of the challenge becomes so large that historically such experimentation has been often not carried out and resulted in significant commercial problems where issues arise in late stage manufacturing development. This latter problem may be further exacerbated by approaches in which final process development is deferred until later clinical trial stages and small process changes may impact on protein stability after significant commercial commitment has been made.

An additional complication is that small-scale assessments, for example processes carried out in microtitre plates at  $\mu\text{l}$  volumes do not impose the same physical stress on the protein materials and as such do not fully indicate the relative performance of molecules in manufacturing type processes. As such there is a conflict between the cost/effort/duration of obtaining test material and the need for representative manufacturing like processes.

This is further complicated by the fact that multiple candidate molecules require assessment for the selection of the most appropriate ones for the manufacturing

platform processes. Additionally there will be multiple downstream process unit operations for which such assessments are required and further, there may exist multiple standard 'platform' manufacturing processes against which assessments are required, thus creating an extensive matrix of testing.

5

Additionally the behaviour of therapeutic proteins may be unpredictable. Such candidate therapeutic proteins may have been modified for multiple purposes; they may be novel fusions of peptide regions never tested before, they may have had their amino acid sequence altered to enhance target binding or to reduce non-specific binding and toxicity issues. This is then to be combined with a range of purification and formulation conditions. As such whilst small-scale e.g. plate based models and in-silico methods can provide guidance the need for practical in-context testing remains.

10

Typical experimental procedures include statistically based DOE (Design of Experiment) procedures in which a test sample is exposed to a process under multiple sets of values for a defined sub-set of operating parameters. Typically a full 2-factorial DOE experiment would involve 16 trials with statistically selected values for 2 operating parameters. The standard approach would be to set-up 16 experiments on the same material which further increases the need for test material. One problem is that the ranges selected for the operating parameter values may be unpredictably inappropriate for the test protein in question and it may become apparent during the experiment that the protein is, for example, unstable in a certain range of values for a particular parameter. As such all the material used in the proportion of the experiments using the aforesaid range is effectively wasted. Although contributing to the knowledge of the performance of the molecule the particular performance boundary may have been identified through the use of less test material.

20

25

Experimental procedures or individual runs can often take a long time to execute as well. For example, the duration of bioreactor runs may be in excess of 14 days and in the case of perfusion bioreactors over 60 days. In these scenarios it can be impossible to predict how the material will perform. This results in a delay in achieving meaningful results and a large amount of valuable material

30

being wasted before a new experiment can be designed. Conventionally, as described above, the only way to speed up development was to run an ever increasing number of parallel experiments using different operating parameters. In practice in order to reduce the material wasted the number of parallel runs is  
5 balanced against the number of re-runs required to produce a meaningful result.

As above, the typical approach to process development is that the scientist uses accumulated knowledge to design a set of parallel experiments exploring multiple operating parameters thought to affect the characteristic(s) for which  
10 there is a desired outcome, for example running a bioreactor at different pH settings with different feeds profiles (e.g. feed type and timing/rate of feed delivery) may affect the product yield obtained. Typically multiple cycles of experimentation are executed with rounds of data analysis and experimental design between them, gradually focusing in on the parameters that impact most  
15 on the desired outcome and refining the values for those parameters.

An example flow diagram illustrating such an approach is shown in Figure 1. The experiment run is first designed manually using statistical design packages (step 10) to create a design for the experimental run (Run Design) (output/input  
20 11). The experimental run is executed (step 12) and data is generated (output 13). The data is analysed manually (step 14) before being stored in a database (step 15). Analysis may of course utilise data already within the database. The results of the data analysis allow the designer to understand the experimental space more completely and to design a second experiment using different  
25 parameters focusing on the new defined area of uncertainty and the experiment process starts again. The entire process is iterative, each iteration leading the designer closer to a more certain solution. The use of data analysis packages to assess such experimental data and design an optimal set of experiments for the next stage is well known, for example, DOE or alternatively (an potentially  
30 including within them DOE) automated experimental design processes also often referred to as 'robot scientist' approaches.

In environments in which, for example, structure function relationships of molecules and synthetic processes are generally less complex than

bioprocessing, such as small molecule synthesis, or for simple bioprocessing unit operations it is known to provide a degree of artificial intelligence, as illustrated in Figure 2. The complexity and variables are reduced, analysis techniques more comprehensive and the data retrieved from a single run can be easily understood and analysed. The parameters for the experiment are first estimated using algorithm based approaches that might include or be considered to be artificial intelligence (AI, step 20) and then the process continues as illustrated in Figure 1. Once the experimental data has been collected, it is analysed using the algorithms/AI (step 24) before being passed to a database for storage and retrieval (step 25) and the results being fed back in to the algorithms/AI to iterate to a new experiment design. Such design does not offer reduction in the volume of material required though it may theoretically reduce the number experimental steps to some degree. Such approaches still required significant volumes of material and place a demand on manual labour.

15

Development of processes and strategies for the manufacturing of biological molecules is complex. Taking the example of the production of therapeutic biological molecules through the culture of microorganisms such as the bacteria E.coli.

20

Firstly there are multiple factors in the creation of such recombinant bacterial strains that add complexity. The details of the DNA constructs carrying the gene for a target protein can have a significant effect on the yield achievable, the growth conditions required to achieve any production and also the nature of the product with respect to stability, structure and function.

25

Additionally there are interactions between such factors, for example it is widely known that the growth conditions, for example temperature at specific stages of the process may affect the yield of functional protein. At higher temperatures such as 37°C, typical growth conditions for E.coli, the organism may produce high yields of protein but due to the fact that recombinant protein sequences are generally derived from different organisms such proteins are not able to correctly form their 3 dimensional structure during the process of translation in the alien host production organism. Reducing the growth temperature and thus reducing

30

the rate of protein production can in some cases enable the correct co-translational protein folding to occur. Thus although the yield of the recombinant protein is less at lower temperature the actual yield of 'functional' protein is higher. The complexity of such systems is further illustrated by for example the fact that the nutrition supplied may limit the availability of certain amino acids essential for the target recombinant protein; this in turn may impact on the rates of production of the target protein, thus there being interactions between nutrition and temperature.

10 These are only two simple factors of many which include; media composition including amino acids, carbon source, micronutrients including metal cofactors and other modulating biological molecules, levels of dissolved gases including oxygen and CO<sub>2</sub>, pH, and osmolality.

15 Whilst such multiple factors are found within chemical synthesis processes the number of potential factors in bioprocessing exceeds that. Additionally different production host strains of E.coli will also respond differently to a given mix of growth conditions and DNA constructs. Searching combinations of DNA construct, host production strain and growth conditions occupies significant amount of time in bioprocessing and though some prediction can be made practical experimentation remains the mainstay of the industry due to the complexity.

25 A further level of complexity found in bioprocessing to a level exceeding that in chemical synthesis arises from the impact of the accumulated history of the production strain. The process used for the generation of sufficient quantity of the production organism to provide an inoculum may have an impact on the performance during the production run.

30 Additionally, history further has an impact during the production bioreactor run (and also in the inoculum preparation culture run). The conditions change during the culture process as biomass increases, waste product levels change and nutrient are consumed amongst other changes. Such changes also potentially

affect the DNA construct itself. As such the complexity due to the number of interacting factors makes practical experimentation essential.

5 In mammalian cell cultures a further challenge arises from the fact that with most current established technologies the insertion of the DNA construct carrying the gene for the recombinant protein may be into the host cell genome may occur randomly. As such the nature of the DNA surrounding the insertion site can have a gross impact on the yield of the protein and the response to the factors described above.

10

There are of course many other factors impacting such processes.

Returning to Figures 1 and 2, the database in which the data is subsequently stored may exist in locally, at an enterprise level or be available more widely through internet based access systems of varying degree of 'openness' commonly referred to as the 'cloud'. In cloud-based experimental design systems, data provided from a series of experimental sources may be curated and made available by different entities and then compared to improve subsequent experiment design.

20

In summary, known bioprocessing development systems require that a full cycle of experimentation be carried out before any refinement for the next set of test conditions are tested, there is an extended duration of development as the experiments iterate towards a conclusion and a very valuable input materials are used inefficiently much of it being wasted.

25

### **Summary of the invention**

The present invention provides for the above problems of material supply to be alleviated. Experiment durations are shortened, data obtained from scarce, valuable materials are maximised and improved results achieved. In some embodiments the invention includes the capability for the scientist to apply bias and priority to such outcomes. To do this, the invention provides systems and methods for automating process development in bioprocessing environments in which the parameters used to execute an experiment run can be modified during

30

the run based on data being generated in parallel in multiple, concurrent experimental systems.

According to a first aspect of the invention there is provided a method for automating process development in a bioprocessing environment. The method comprises: executing a first experiment run according to a set of parameters; retrieving a second real-time set of data of an experiment run being executed in parallel, analysing the retrieved first real-time set of data and the second real-time set of data to determine an adjusted set of parameters; and, modifying, based on the analysis, the parameters upon which the experiment run is being executed during execution of the run such that the run continues to be executed according to the modified set of parameters.

Thus, the invention directly leads to an increase yield of data and knowledge of the experimental space with thorough exploitation of all materials utilised in the execution of the run. This leads to significantly increased definition of the experimental space from a single experimental run and hence the reduction in the number of rounds of experimentation required and a resulting decrease in development cycles, associated reduction in the time consuming preparation of high value materials and an overall reduction in development cycle time in the industry with benefits for time to market, development costs and resources.

Additional benefits include the increased definition of the functional operating space for particular candidate therapeutics which increases the likelihood of scale-up of such processes to manufacturing scale and reduces the significant cost and commercial impact of delays late in development such as delayed launches, regulatory review issues, etc. Experimental parameters can be refined multiple times within a single set of experimental runs. Such refinement can occur on a dynamic basis i.e. as a measured parameter in the experiment reaches a pre-set boundary set, for example describing the target production process control parameters can be adjusted based on the dataset from the parallel experiments in process to adjust the control parameters and adjust the trajectory of the model process experiment in question. Thus such a system could be used to execute targeted process development in which the

experimental system dynamically identifies a series of parameter settings that can be applied to meet a target goal. Such a series of parameter settings, potentially differing, might be generated in all the parallel experiments providing alternative control strategies. Such systems provide clear understanding of the boundaries of a region of poor performance for a particular candidate therapeutic protein to become apparent in the experiment allowing the experiments to be re-tasked to obtain a different result or examine a different condition or criteria.

Benefits may be made possible by interconnected units and parallel experimentation and systems which are connected to share data. Through the parallel monitoring of the data, such as protein status, using in/on-line analysis, more useful sets of parameter values can be automatically identified by the system through combined analysis of all of the experiments underway and the experiments in the problem area 're-tasked' through the automatic adjustment of the experiment values. If conditions are identified in one experiment, another experiment can be modified to ensure the condition does not occur or parameters adjusted to ensure that the experiment is re-tasked for a different purpose based on the analysis of the other experiment.

For example, if one experiment within a parallel run shows increased utilisation of a valuable material, the conditions of the other experiment can be adjusted so that the utilisation is tested in the other environment or the environment amended to the same to achieve that same high utilisation.

The method allows for full exploitation of extended processes such as perfusion allowing multiple experiments to be carried out in combination with automated process/experiment design.

The analysis step may include comparing the retrieved first real-time set of data against a predetermined set of criteria and in which the parameters are modified based on the comparison. The criteria may be set automatically or by an operator. In this way, for example, the method is able to determine at which point the experiment should be modified or indeed when.

Furthermore the method may further comprise: retrieving a historical set of data of an experiment run previously executed, wherein the analysing step is performed based on the historical set of data. In this way experiments can be re-tasked to revisit an earlier set of parameters to repeat and confirm previous findings or explore areas of variation. Moreover trends in previous experiments can be avoided or used to improve the current experiment. For example if previous experiments tended in a particular direction the current experiment can be modified if the same trend is detected.

Optionally the method may comprise retrieving a third-party set of data of an experiment run executed by a third party. The third-party set of data may be retrieved from a cloud computing system or from a local database. The analysing step may be performed based on the third-party set of data. Thus trends or experience from other experiments run by others can be utilised to improve the current experiment.

The set of data may include one or more selected from a group comprising yield, protein conformation, protein function, protein concentration, viscosity, derived pH, conductivity, media composition, gas utilisation, cell health and status parameters. The parameters may be measured directly or indirectly during the execution of the experiment.

The parameters may include one or more selected from a group comprising: pH, media composition, temperature, gas supplies, pressure, flow rates and mixing criteria. Many other parameters exist that may be of importance and could be identified by expert in the art and will be identified in the future, for example specific micro nutrients

The method may further comprise retrieving a target outcome and a set of boundary conditions, wherein the analysing step is performed based on the retrieved target outcome and set of boundary conditions. The automation is thus controlled within a defined set of limits and towards a particular goal or outcome. The conditions and outcome may be set by the operator. For example, yield of functional protein may be maximised through one or more series of parameter

changes such that the parameter values used remain within defined pre-set boundaries, thus automatically creating a manufacturing process that fits one or more of a pre-existing set of 'platform' manufacturing processes.

- 5 The analysis step may comprise extrapolating the experiment to predict experiment data. In this way the art is improved since the extrapolation may occur from real data and actual samples rather than only extrapolating from related materials. For example, data never tested before in this fashion can be extrapolated more accurately as more data is retrieved. Moreover, in some  
10 embodiments data from parallel runs can be used to improve the extrapolation.

According to a second aspect of the invention, there is provided a system for automating process development in a bioprocessing environment, the system comprising: an executing unit configured to execute a first experiment run  
15 according to a set of parameters; and, an analysis unit configured to: retrieve a first real-time set of data of the experiment run while the experiment run is being executed; and, analyse the retrieved first real-time set of data to determine an adjusted set of parameters, wherein the executing unit is further configured to modify, based on the analysis, the parameters upon which the experiment run is  
20 being executed during execution of the run such that the run continues to be executed according to the modified set of parameters.

The analysis unit may be further configured to compare the retrieved first real-time set of data against a predetermined set of criteria and in which the  
25 parameters are modified based on the comparison. The analysis unit may be further configured to: retrieve a second real-time set of data of an experiment run being executed in parallel, wherein the analysis is performed based on the second real-time set of data. The system may include an interconnected control units and a database which may be local to the units. The sets of data may be  
30 stored in a database which may be in the cloud, remotely or local to the parallel experiments.

The analysis unit may be further configured to: retrieve a historical set of data of an experiment run previously executed, wherein the analysis is performed based

on the historical set of data. The analysis unit may be further configured to: retrieve a third-party set of data of an experiment run executed by a third party, the third-party set of data being retrieved from a cloud computing system, wherein the analysis is performed based on the third-party set of data. In theory  
5 any remote computing system may be used. The set of measured data collected dependant on the system and process under study may include one or more selected from a group including yield, protein conformation, protein function, protein concentration, viscosity, derived pH, conductivity, media composition, gas utilisation, cell health and status parameters. The parameters  
10 may be measured directly or indirectly during the execution of the experiment. The parameters may include one or more selected from a group comprising: pH, media composition, temperature, gas supplies, pressure, flow rates and mixing criteria. The parameters may be parameters controlling a run dependant on the system and process under study. The analysis unit may be further configured  
15 to: retrieve a target outcome and a set of boundary conditions, wherein the analysis is performed based on the retrieved target outcome and set of boundary conditions. The analysis unit may be further configured to: extrapolate the experiment to predict experiment data.

20 According to a third aspect of the invention a computer program is provided which, when executed by a processor, causes the processor to carry out a method according to any of the above aspects. A computer-readable medium storing a computer program may also be provided.

## 25 **Detailed Description**

Examples of systems and methods in accordance with the invention will now be described with reference to the accompanying drawings, in which:-

30 Figure 1 shows an experiment design approach in known bioprocessing development;

Figure 2 shows a known design approach in chemistry synthesis;

Figure 3 shows a flow diagram of a first embodiment of the present invention;

Figure 4 demonstrates an exemplary architecture in which an embodiment of the present invention can be performed;

Figure 5 shows an iterative process according to an embodiment of the present invention for identifying optimal experiment parameters; and,

Figure 6 shows an example set of data from parallel experiment runs of a hypothetical experiment.

5

The following are examples of systems and methods for dynamic experimental design and execution. An artificial intelligence based method is provided for monitoring multi-experiment designs dynamically during an experiment as illustrated in Figure 1. The present invention may be embodied in a control system, automation unit, measurement devices, scheduler, software package operating on a computer with necessary inputs or on an artificial intelligence software package. Each term may be used interchangeably to refer to a computer unit connected to data outputs of one or more experiments.

10

15 Optionally the process begins with a software package designing a set of parameters for a particular experiment or set of parallel experiments (step 30). The parameters may be adjusted manually by an operator or of course may be entirely set manually according to the experience of the scientist. The experiment then proceeds to be run (step 11) with manual initiation or may be initiated in a fully automated fashion in a system with appropriate robotics. In the known manner, the run is executed in an automated fashion (step 12). During execution, data is extracted from the experiment (step 31) and is analysed by the artificial intelligence or software package (step 32). A modified experiment design is determined (step 33) and the experiment continued based on the modified experiment design and set of parameters chosen for that experiment. The cycle of analysis and adjustment may be at fixed time intervals or 'continuous'

20

25

30

Embodiments of the invention provide for the development of systems capable of executing multiple complex processes in parallel under the control of a single system. Data may be used from one such experiment to adjust the parameter settings of an experiment being carried out in parallel in a dynamic fashion effectively creating a large number of dynamically adjusted experiments utilising

data generated from each other in combination with existing data that may be present in a number of forms and virtual locations.

A simplified hypothetical example will now be described. The pH setting for six  
5 of the bioreactors in an experimental run is 7.0. For another set of six bioreactors in the same run it is 7.5. During the experiment on-line real time analysis indicates that the lower pH is a principal component of lower productivity. In response to this the system adjusts the pH to a higher value to enable increased depth of analysis of other parameters (or the value of a pH  
10 step) in achieving the goal of higher productivity.

In addition to the above singular process, the algorithms comprising the process may extract data from historic databases and combine this with dynamic real time data of the experiments being run.

15

One advantage of the dynamic design approach is that as the boundary of the poor performance region becomes apparent in the experiment. Through the parallel monitoring of the protein status using in/on-line analysis systems, more useful sets of parameter values can be automatically identified by the system  
20 through combined analysis of all of the experiments underway and the experiments in the problem area 're-tasked' through the automatic adjustment of the experiment values.

Such re-tasking during the experiment can be utilised in a number of additional  
25 ways, for example in a system that automatically runs the component of experiments of a DOE in a sequential manner the parameter sets for subsequent experiments can be based on those of experiments within the automated run in a fully automated fashion.

30 Additionally, the utilisation of different statistical approaches to identify optimal operating conditions and performance plateaus for operating conditions such as 'hill climb algorithms'.

The dynamic design approach also allows an experiment to be re-tasked to revisit/replicate an earlier set of experimental parameters to repeat and confirm previous findings within the experiment to explore levels of variation or to understand changes within the condition of a key part of the experimental system, such as the development or depletion of a gel layer on the filter in a UF/DF system.

It has previously been described that the dynamic design approach has particular utility in parallel experimental runs where experiments are designed using different conditions and run in parallel to provide a series of data in a short time. In extended experimental runs such as perfusion cell culture where stable conditions can be established, the operator may set the parameter conditions, allow the running bioreactor to become stable and then follow a read-out for a period of stable operation. The operator would then adjust the operating parameter settings, allow the system to stabilise and then again follow the readout. The invention is particularly powerful in such cases in that firstly it allows the automation of such systems and the  $n+1$  set of parameter values can automatically be based on those from experiment  $n$ . Even more powerful is the application of such approaches in parallel systems, for example a multiple parallel bioreactor system capable of running many sequential experiments. Data from all the experiments under execution within the parallel system run can be combined to define the parameter values for the next stage of the experiment in any one bioreactor.

Also in such cases there is no requirement of the completion of a particular experimental phase before re-tasking based on other data collected in the system occurs.

Re-tasking and calculation of new parameter values during the run may involve machine learning or other forms of AI and existing local and remote data sets.

Referring once again to Figure 3, at the conclusion of the experiment, the data is extracted and analysed (steps 13 and 36) before a new experiment or set of parameters is designed and the process is iterated. The data analysis, modified

in-experiment design and during run data may all be stored in a database (step 37). The control system may identify a next step set of experiments to be executed and simply issue an instruction for request that a technician to resupply the system with a set of defined consumable components (where these are not simply configured for automated supply to the system), and to confirm completion of loading. After which the system would execute the run (step 11).

The systems may generate novel additional data and the addition of that to a database may increase the discriminatory power of the historic database (step 37). The enhanced information content of this database will increase the success and speed of such a system.

An exemplary architecture in which the above may be embodied is illustrated in Figure 4. System components are illustrated, in for example a bioreactor system. The artificial intelligence scheduler is shown as item 40 and in communication with the other elements of the system. The scheduler will first provide input into the experiment system 41. The experiment system generates system data measured or derived directly on the experimental system in real time which may be transferred into and/or via a local controller which additionally provides control signals for the experiment. System data is output 43 together with data measured indirectly through the processing of samples of the experimental system external to the system and analysed to generate data 42 relating to the experiment status such as protein quality or yield, to a database which is itself in communication with the scheduler. That is, the scheduler retrieves data of the experiment from a database 44. Optionally of course it may retrieve this data directly.

The scheduler may receive inputs from the operator to aid in defining the experiment parameters. These may be goals, boundaries or other knowledge the scheduler is not aware of but the operator may have that may aid in defining the experiment.

The scheduler 40 may retrieve data from and upload data to the cloud, illustrated here as collaboration data. In this way the scheduler may learn from experiments run by third parties and may help others for mutual benefit.

5 A key part of the architecture is an applications database 46 with which the scheduler 40 is in communication. The applications database 46 typically might contain information on the operation of such systems and requirements of known and presumed applications and may aid in defining the experiment and may include parameter relationships such as rpm to KLa etc., operating boundaries  
10 for the experiments as well as scalability algorithms to aid in the parameter modification.

Tools 47 may be used by the scheduler such as DOE tools which may aid in the experimentation process. Optionally, the scheduler may include a series of  
15 APIS to enable interface with third party tools.

It will be clear to the skilled person that the database and scheduler may be comprised in one system where appropriate or may be made up of multiple functionally separated blocks.

20

The flow charts and descriptions thereof should not be understood to prescribe a fixed order of performing the method steps described therein. Rather, the method steps may be performed in any order that is practicable. Methods and processes described herein can be embodied as code (e.g., software code)  
25 and/or data. Such code and data can be stored on one or more computer-readable media, which may include any device or medium that can store code and/or data for use by a computer system. When a computer system reads and executes the code and/or data stored on a computer-readable medium, the computer system performs the methods and processes embodied as data  
30 structures and code stored within the computer-readable storage medium. In certain embodiments, one or more of the steps of the methods and processes described herein can be performed by a processor (e.g., a processor of a computer system or data storage system). It should be appreciated by those skilled in the art that computer-readable media include removable and non-

removable structures/devices that can be used for storage of information, such as computer-readable instructions, data structures, program modules, and other data used by a computing system/environment. A computer-readable medium includes, but is not limited to, volatile memory such as random access memories (RAM, DRAM, SRAM); and non-volatile memory such as flash memory, various read-only-memories (ROM, PROM, EPROM, EEPROM), magnetic and ferromagnetic/ferroelectric memories (MRAM, FeRAM), and magnetic and optical storage devices (hard drives, magnetic tape, CDs, DVDs); network devices; or other media now known or later developed that is capable of storing computer-readable information/data. Computer-readable media should not be construed or interpreted to include any propagating signals.

One example to which the invention might be applied is the automated optimisation of mammalian culture parameters in a bioreactor in order to achieve maximum yield.

The interest of the user is in identifying the optimal conditions providing the highest yield of a recombinant protein produced by the mammalian cell line being cultured. Key parameters known generally to have an effect on the yield of recombinant CHO strains in bioreactors are varied in experiments designed using statistical principles (Design Of Experiments (DOE)) to identify the relative effect of each of the factors being varied. An experiment involving two factors being varied typically involves 16 individual experiments in one run of an automated parallel bioreactor system such as the Sartorius ambr250HT system capable of supporting 24 bioreactors in parallel. pH and temperature are known to impact the yield of functional protein and the range explored for such values in culture is shown in Table 1 with low, medium and high values for each parameter being defined

Table 1. Parameter value ranges for testing

Level	L	M	H
pH	6.5	7.0	7.5
Temp	32	35	38

Using a Central Composite Face (CCF) Design of Experiment with 4 centre points, 12 individual experimental runs are defined as shown in Table 2 to explore the significance of each factor in determining yield and the interactions between those factors.

5

Table 2

Experiment	pH	Temp
1	6.5	32
2	6.5	35
3	6.5	38
4	7.0	38
5	7.5	38
6	7.5	35
7	7.5	32
8	7.0	32
9	7.0	35
10	7.0	35
11	7.0	35
12	7.0	35

- 10 The bioreactors are set with the appropriate parameter set-points up by the operator and the experiment initiated with these conditions. The production of functional recombinant protein is measured on a regular basis using either an in-line assay system (such as Ramen, NIR) or regular manual or automated sampling, for example in the widely used Sartorius ambr250 or ambr15 systems,
- 15 to either an integrated at-line or an off-line analyser (such as Pall fortébio Octet system, Nova Biomedical Bioprofile Flex system) and the data recorded directly by the system or transferred manually or directly electronically to the system.

- 20 Following a certain duration of culture it becomes clear that Experiments 1 to 3 do not yield any protein and the associated algorithm applied to the data

identifies that such effect is significantly associated with the pH of the bioreactor. It is determined by the algorithm that continuing such experiments at pH<7.0 would be of no value beyond establishing that protein yield at pH 6.5 is insignificant. The system is programmed so that in such a case the pH is automatically adjusted based on the pH of bioreactors in the same run known to be expressing protein at high levels. For example, in the case that the highest yield of protein in the parallel system is from bioreactors with pH set-points of pH7.5, the pH highest being tested, adjustment of the pH of the bioreactors 1 to 3 to pHs of 7.8, higher than that that of the high producing bioreactors, allows the upper limit of pH for yield to be explored in the same experiment utilising resources that would otherwise have been wasted. Therefore, an additional round of experimentation may be negated or reduced in scale.

Such approaches could be applied to a number of bioprocessing criteria relating to primary production in microbial or cell cultures, for example protein quality traits or cell specific productivity (the production of target protein per cell per unit time, typically pg/cell/day). Equally such approaches might be applied to downstream processing applications, for example the optimisation of pressure and flow parameters in ultrafiltration and diafiltration to optimise flux rates.

A number of strategies may be applied to take advantage of dynamic data generated in parallel automated systems. A further example is provided in below and illustrated in Figure 6.

Six bioreactors are set-up and run initiated with the same putative production cell line and the productivity of the cells in each bioreactor is calculated and monitored over time. At specific time points, (the frequency of which might vary depending on the traits being measured and the analysis methodology used) an algorithm is applied to the data and based on specific settings for parameters in the algorithm the operating set points for a subset of the bioreactors is altered. For example in Figure 6, the operating parameter set points are adjusted for the two lowest performing bioreactors, either to a pre-set formula or based upon the data from within those bioreactors and the other bioreactors within the run.

The performance of the bioreactors for the trait is then measured further and at time point 2 the two lowest performing bioreactors are adjusted. The process is then repeated allowing identification of more optimal conditions for the trait. Depending on the detailed context of the experiment additional factors may need  
5 to be considered but the principle remains the same.

A number of variants of the approach are possible including:

- The design of any algorithm may be varied extensively with respect to the data considered in the performance assessment, the set of bioreactors from  
10 which data is considered, the subset of bioreactors on which parameters are altered, the strategy of the alteration i.e. to make the adjusted bioreactors more or less like those on which the performance metric has been measure as more optimal or distinct in some other dimension.
- The initial set of parameter settings for the bioreactors may be based on  
15 DOE whereas subsequent adjusted parameter setting are not or subsequent parameter sets may be based on DOE.
- Data from within each bioreactor may be utilised differently or weighted differently to other bioreactors from within the set, equally other individual bioreactors within the set may be weighted differently based on their current or  
20 historic performance.
- Data from sources external to the system and/or current may be utilised including data from local knowledge databases based on previously acquired data, enterprise level databases, external private and public data bases and also through dynamic requests for expert person input.

25

Using data collation from parallel runs a number of approaches to automated process development are possible.

In Figure 6 it could for example be identified that the total accumulated product  
30 in reactor 5 was higher than any other bioreactor or this might be combined with other parameters such as activity, or % of molecules a specific structural form. This may be the result of specific physiological events triggered by the variation in bioreactor conditions providing the appropriate environment optimal for the trait that would not have otherwise been identified due to the complex nature of

the biology. The series of parameter settings for reactor 5 may then be taken as the most optimal process for manufacture and transferred to larger production scale systems or used as the basis for further process development or research. Algorithms may be specifically designed for such automated process development.

Algorithms may be designed for such experiments including boundary conditions for the traits being assessed. For example, in the above experiment, it is generally known that cell productivities of 20 pg/cell/h or higher are achievable in mammalian cultures. The algorithms may be designed such that if the productivity in a particular reactor is below a set threshold value or some function of that of other reactors in the same run the parameter settings may be adjusted to a value predicted by the system to provide more yield based on the parameter settings being used with high yielding and low yielding bioreactors in the same run. Operational settings in such algorithms could be designed to cluster around parameters known to be high yielding or to explore other distinct regions of experimental space or both.

Such algorithms may use both data from other bioreactors running concurrently in the same system, historical data held within the system or in an enterprise database or third party data.

Further boundary values for operational parameters in such algorithms may be set such that they relate to one or more of a set of platform manufacturing processes used by the manufacturing company. As such any such automatically derived process would effectively optimised to operate within those boundaries and therefore be directly applicable to the manufacturers standard operating processes.

Other strategies enabled by parallel automated systems include the delay of the progression of certain experimental units within the parallel system relative to others; for example delaying the start of bioreactors 4-6 until after bioreactors 1-3 would allow data on the start-up phase from the first set of bioreactors to be used to optimise the start-up phase for the second set of bioreactors. Such

delays might easily be programmed into an automated system and combined with the aforementioned approaches. Approaches used might include automated reduction in the temperature of certain bioreactors in the run to delay the progression of those cultures until a specific time point.

5

Additionally, strategies include distribution of one experimental unit to many; for example, samples from one bioreactor of a set of parallel bioreactors that is identified as having a specified performance (highest, lowest etc.) for a particular trait might be used to inoculate other parallel, but as yet unseeded bioreactors in the same system. For example, the first set of bioreactors from which the chosen bioreactor is selected might contain different cell lines, the most effective of these then being selected for process development

Above, processes have been described in which experiments can be designed to be dynamically assessed and individual experiments 're-tasked' to explore new parameter ranges once the value of a first set of parameter conditions has been evaluated against set measures including other experimental parameter sets running in the same parallel experimental system. The dynamic design enables a user to: load a 'target outcome' and boundary conditions for the process; be free of intensive interaction with the system other than loading supplies and removing waste; and, receive a final process design optimal for a scaled-up reactor system with reduced time and cost of materials. The approach has particular value in that decisions on optimal experimental design can come once the experiment has started and also be implemented before the next round of experimentation – hence the number of rounds of evaluation required is reduced. Such approaches are beneficial due to the complexity of such biological systems and limited capability to predict the outcome, particularly in the early stages of developments when working with new and novel proteins. Further, full automation of sequential experimentation with automated evolution of the experimental design but within a single experimental run is allowed and finally but not exclusively, the approach allows for the use of experimental data from parallel channels in the same experimental system.

The dynamic design approach for process development for a broad range of applications, including but not limited to: selection of biological products for therapeutic and industrial purposes; upstream biological process development e.g. for biological pharmaceutical production, industrial product fermentation, cell based therapy processes etc.; downstream process development e.g. for biopharmaceutical product purification, industrial bio-product purification; formulation of biological molecules for therapy; and. non-biological applications in which parallel experimentations and/or sequential stages of experimentation are used for process development.

10

It can therefore be seen that the dynamic design approach described provide key benefits such as: reduced development cycles; increased efficiency of use of limited resources with significant added value; reduced costs through decreased labour, facilities time, resources; decreased project management complexity; and, increased power and significance of results.

15

CLAIMS

1. A method for automating process development in a bioprocessing environment, the method comprising:
- 5           executing a first experiment run according to a set of parameters;  
          retrieving a first real-time set of data of the experiment run while the experiment run is being executed;  
          retrieving a second real-time set of data of an experiment run being executed in parallel,
- 10           analysing the retrieved first real-time set of data and the second real-time set of data to determine an adjusted set of parameters; and,  
          modifying, based on the analysis, the parameters upon which the experiment run is being executed during execution of the run such that the run continues to be executed according to the modified set of parameters.
- 15
2. A method according to claim 1, in which the analysis step includes comparing the retrieved data against a predetermined set of criteria and in which the parameters are modified based on the comparison.
- 20
3. A method according to claim 1 or 2, in which the method further comprises:
- retrieving a historical set of data of an experiment run previously executed,  
          wherein the analysing step is performed based on the historical set of
- 25           data.
4. A method according to any preceding claim, in which the method further comprises:
- retrieving a third-party set of data of an experiment run executed by a
- 30           third party, the third-party set of data being retrieved from a cloud computing system,  
          wherein the analysing step is performed based on the third-party set of data.

5. A method according to any preceding claim, in which the set of data includes one or more selected from a group comprising yield, protein conformation, protein function, protein concentration, viscosity, derived pH, conductivity, media composition, gas utilisation, cell health and status parameters.
6. A method according to any preceding claim, in which the parameters include one or more selected from a group comprising: pH, media composition, temperature, gas supplies, pressure, flow rates and mixing criteria.
7. A method according to any preceding claim, further comprising: retrieving a target outcome and a set of boundary conditions, wherein the analysing step is performed based on the retrieved target outcome and set of boundary conditions.
8. A method according to any preceding claim, in which the analysis step comprises extrapolating the experiment to predict experiment data.
9. A system for automating process development in a bioprocessing environment, the system comprising:  
an executing unit configured to execute a first experiment run according to a set of parameters; and,  
an analysis unit configured to:  
retrieve a first real-time set of data of the experiment run while the experiment run is being executed;  
retrieve a second real-time set of data of an experiment run being executed in parallel; and,  
analyse the retrieved first real-time set of data and the retrieved second real-time set of data to determine an adjusted set of parameters, wherein the executing unit is further configured to modify, based on the analysis, the parameters upon which the experiment run is being executed during execution of the run such that the run continues to be executed according to the modified set of parameters.

10. A system according to claim 9, in which the analysis unit is further configured to compare the retrieved data against a predetermined set of criteria and in which the parameters are modified based on the comparison.

5

11. A system according to claim 9 or 10, in which the analysis unit is further configured to:

retrieve a historical set of data of an experiment run previously executed, wherein the analysis is performed based on the historical set of data.

10

12. A system according to any of claims 9 to 11, in which the analysis unit is further configured to:

retrieve a third-party set of data of an experiment run executed by a third party, the third-party set of data being retrieved from a cloud computing system,

15

wherein the analysis is performed based on the third-party set of data.

13. A system according to any of claims 9 to 12, in which the set of data includes one or more selected from a group comprising yield, protein conformation, protein function, protein concentration, viscosity, derived pH, conductivity, media composition, gas utilisation, cell health and status parameters.

20

14. A system according to any of claims 9 to 13, in which the parameters include one or more selected from a group comprising: pH, media composition, temperature, gas supplies, pressure, flow rates and mixing criteria.

25

15. A system according to any of claims 9 to 14, in which the analysis unit is further configured to: retrieve a target outcome and a set of boundary conditions,

wherein the analysis is performed based on the retrieved target outcome

30

and set of boundary conditions.

16. A system according to any of claims 9 to 14, in which the analysis unit is further configured to: extrapolate the experiment to predict experiment data.

17. A computer program which, when executed by a processor, causes the processor to carry out a method according to any one of claims 1 to 8.
  
  18. A computer-readable medium storing a computer program according to claim 17.
- 5

Fig. 1

(Prior Art)

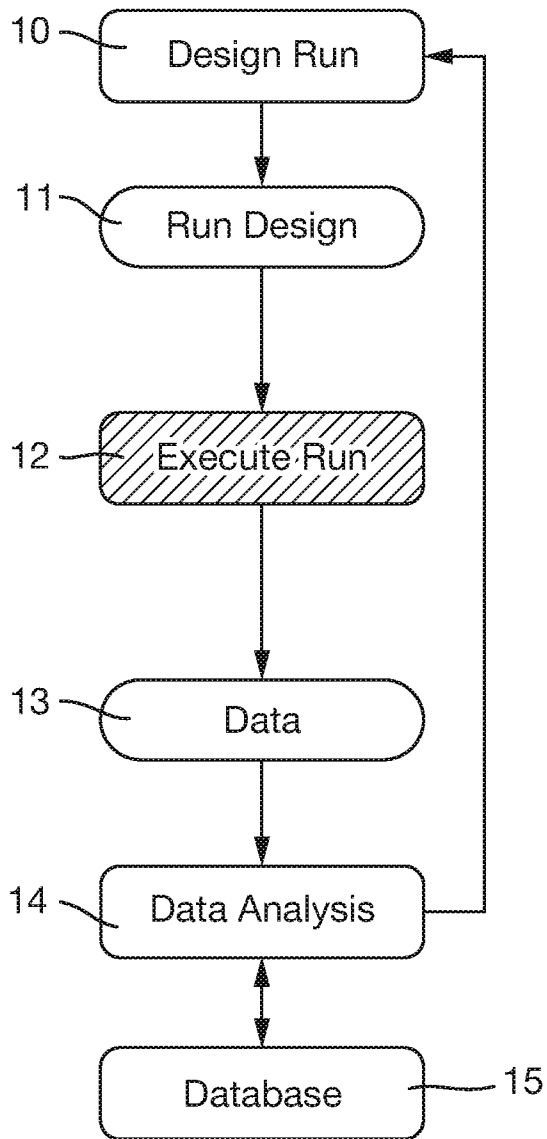


Fig. 2

(Prior Art)

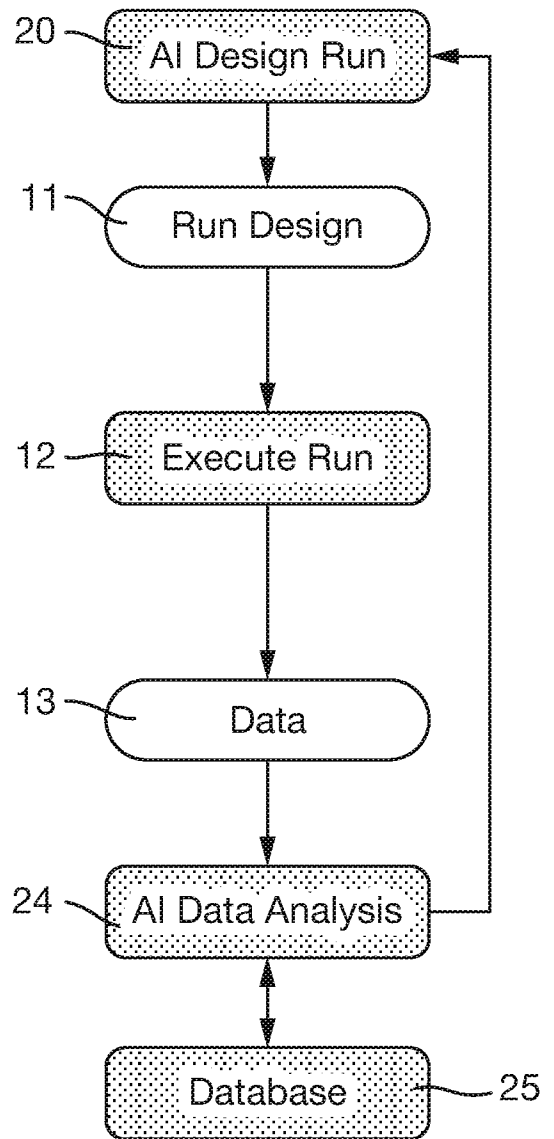


Fig. 3

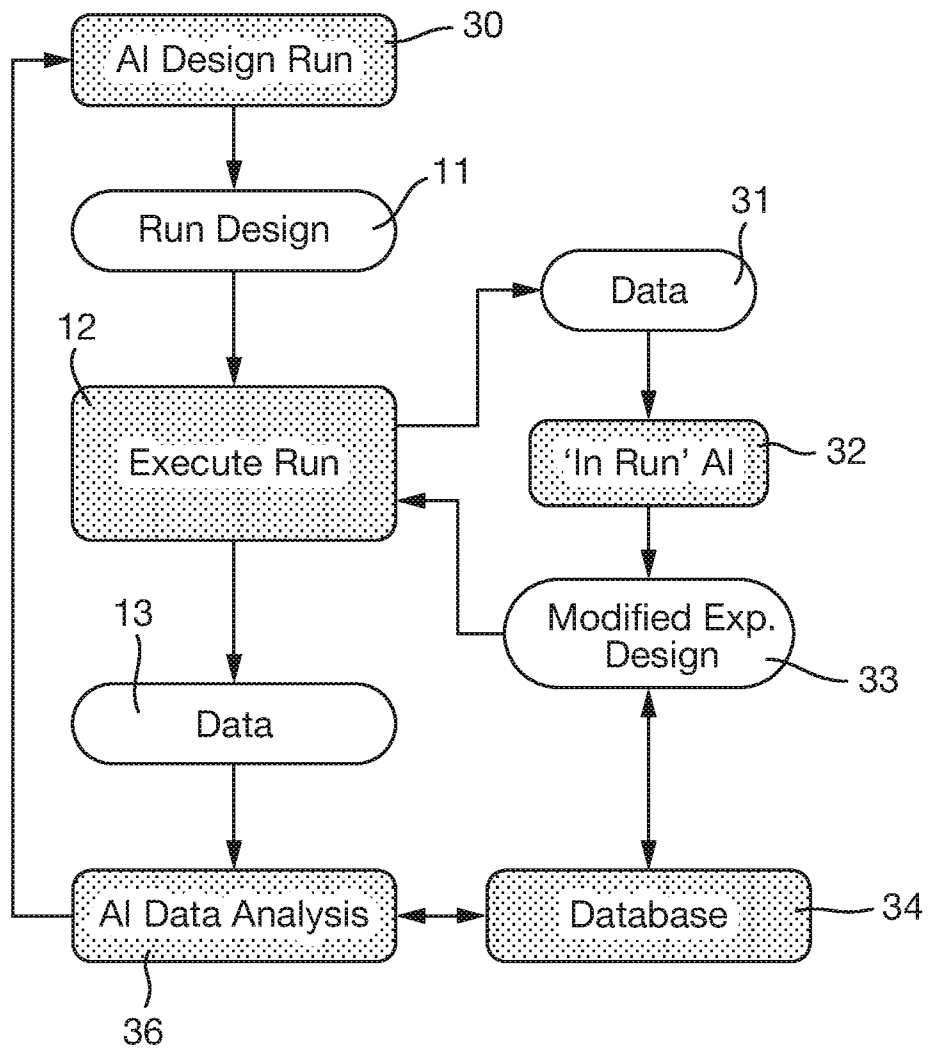


Fig. 4

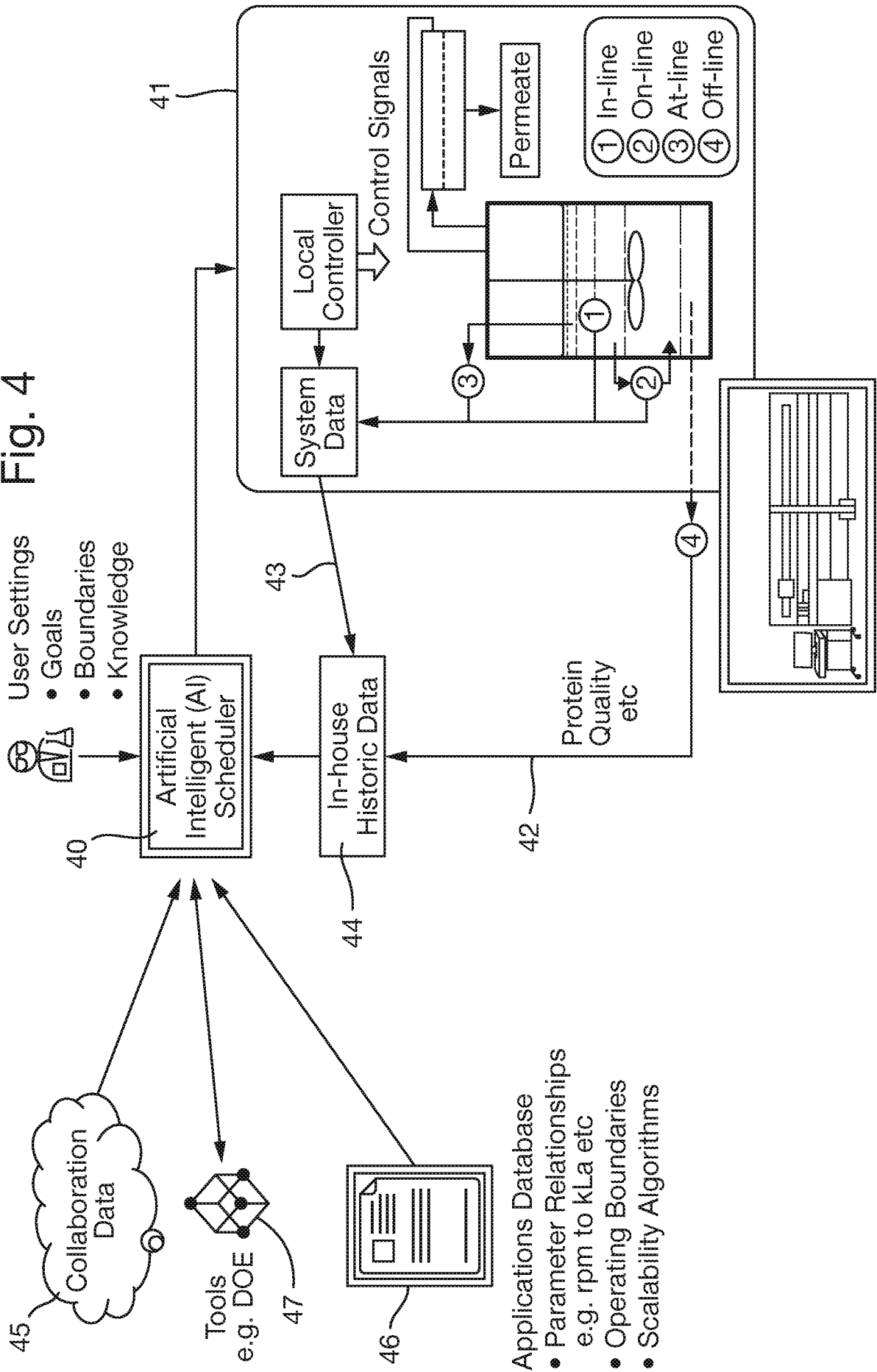


Fig. 5

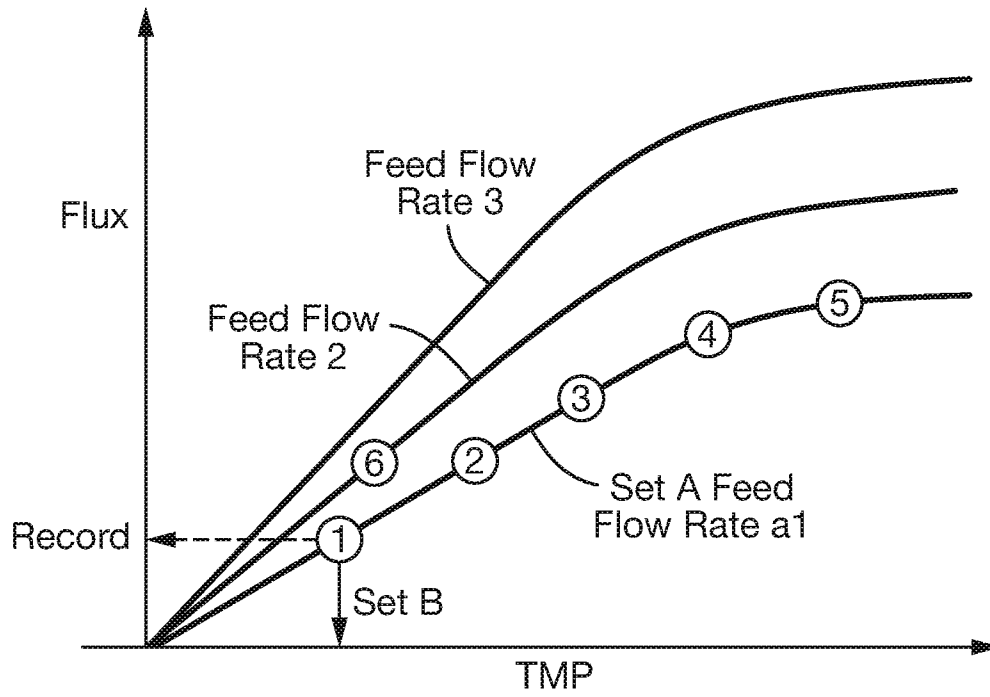
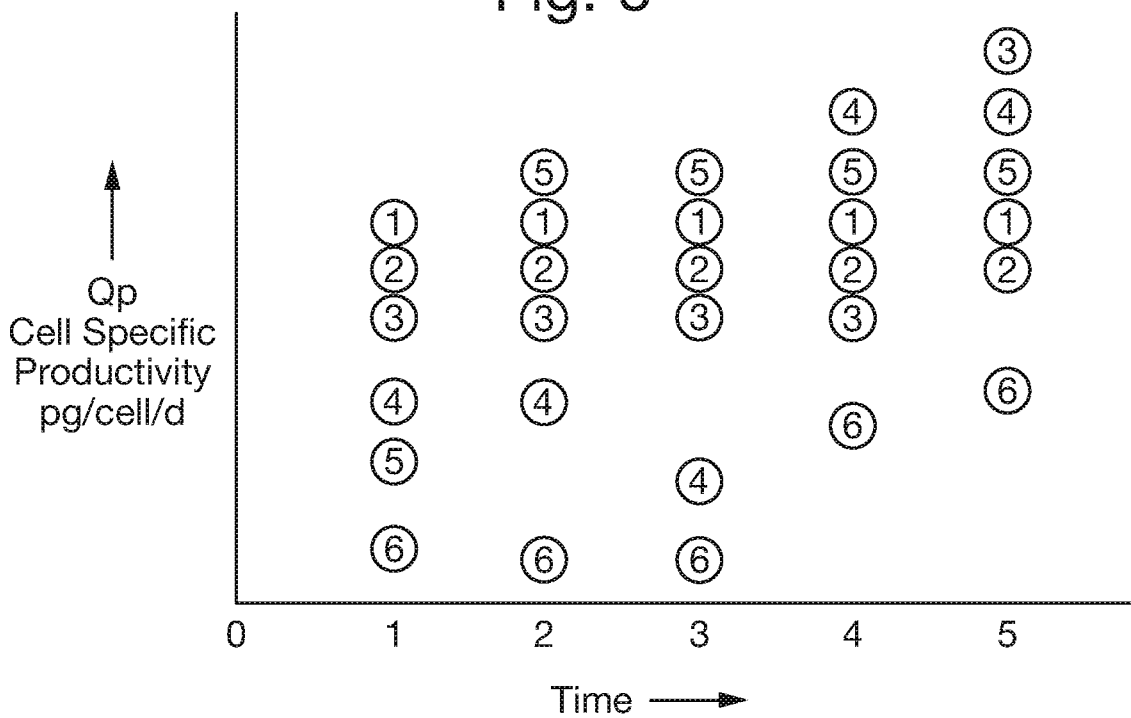


Fig. 6



**INTERNATIONAL SEARCH REPORT**

International application No  
PCT/GB2017/051326

**A. CLASSIFICATION OF SUBJECT MATTER**  
 INV. G01N31/00 G05B13/02 G05B15/02 G06F19/00  
 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)  
 G01N G05B G06F

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.
X	US 2008/252886 A1 (BODENSTAFF EMILIO RENE [NL] ET AL) 16 October 2008 (2008-10-16)	1-3, 5-11, 13-18
Y	paragraph [0008] paragraph [0054] paragraph [0004] paragraph [0023] paragraph [0042] paragraph [0050] - paragraph [0051]	4,12
Y	US 2010/250147 A1 (VERSEPUT RICHARD P [US]) 30 September 2010 (2010-09-30) paragraph [0087] paragraph [0181]	4,12
A	WO 02/093297 A2 (TRANSFORM PHARMACEUTICALS INC [US]) 21 November 2002 (2002-11-21) figures 1,5	1,9

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 <b>5 July 2017</b>	Date of mailing of the international search report <b>13/07/2017</b>
---	---

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

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2017/051326

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2008252886	A1	16-10-2008	CA 2623216 A1 29-03-2007
			CN 101273264 A 24-09-2008
			EP 1934587 A1 25-06-2008
			JP 2009509162 A 05-03-2009
			KR 20080075096 A 14-08-2008
			US 2008252886 A1 16-10-2008
			WO 2007035087 A1 29-03-2007
-----			
US 2010250147	A1	30-09-2010	US 2010250147 A1 30-09-2010
			US 2013013221 A1 10-01-2013
-----			
WO 02093297	A2	21-11-2002	AU 2002303683 A1 25-11-2002
			CA 2447047 A1 21-11-2002
			EP 1395808 A2 10-03-2004
			WO 02093297 A2 21-11-2002
-----			