2. Tewari, H. C., Dixit, M. M., Madhava Rao, N., Venkateshwarlu, N. and Vijaya Rao, V., Crustal thickening under the Palaeo-meso-Proterozoic Delhi Fold Belt in Northwestern India: evidence from deep reflection profiling. *Geophys. J. Int.*, 1997, **129**, 657–668.
3. Tewari, H. C. *et al.*, Deep crustal reflection studies across the Delhi–Aravalli fold belt: Results from the northwestern Part – by Controlled Source Seismic Group. *Geological Society of India, Memoir*. 1995, vol. 31, pp. 383–402.
4. Prasad, B. R., Tewari, H. C., Vijaya Rao, V., Dixit, M. M. and Reddy, P. R., Structure and tectonics of Proterozoic Aravalli–Delhi Fold Belt in north western India from deep seismic reflection studies. *Tectonophysics*, 1997, **288**, 31–41.
5. Reddy, P. R. *et al.*, Deep seismic reflection profiling along Nandsi–Kunjer section of Nagaur–Jhalawar transect: Preliminary results, In Continental Crust of Northwestern and Central India. *Geological Society of India. Memoir*, 1995, **31**, 353–372.
6. Reddy, P. R. and Vijaya Rao, V., Structure and tectonics of Indian peninsular shield – Evidences from seismic velocities. *Curr. Sci.*, 2000, **78**, 899–906.
7. Vijaya Rao, V., Rajendra Prasad, B., Reddy, P. R., Tewari, H. C., Evolution of the Proterozoic Aravalli–Delhi Fold Belt in northwestern Indian Shield from seismic studies. *Tectonophysics*, 2000, **327**, 109–130.
8. Satyavani, N., Dixit, M. M. and Reddy, P. R., Crustal velocity structure in Nagaur–Rian sector of Aravalli Fold Belt, India, by using reflection data. *J. Geodyn.*, 2001, **31**, 429–443.
9. Sinha-Roy, S., In Precambrian of the Aravalli Mountain Rajasthan, India. *Mem. Geol. Soc. India*, 1988, **7**, 95–108.
10. Sinha-Roy, S., Malhotra, G. and Guha, D. S., A transect across Rajasthan Precambrian terrain in relation to geology, tectonics and crustal evolution of south central Rajasthan. In *Continental Crust of North Western and Central India* (eds Sinha-Roy, S. and Gupta, K. R.), *Mem. Geol. Soc. India*, 1995, **31**, 63–90.
11. Megallaa, M., Depth conversion by the use of forward and inverse ray-path modelling and compensation for the heterogeneity of the Earth. *Explor. Geophys.*, 1989, **20**, 445–467.
12. Heron, A. M., The geology of central Rajputana. *Mem. Geol. Soc. India*, 1953, **79**, 1–389.
13. Sychanthavong, S. P. and Desai, S. D., Proto plate tectonics controlling Precambrian deformations and metallogenetic epochs of northwestern peninsular India. *Miner. Sci. Eng.*, 1977, **9**, 218–236.
14. Sinha-Roy, S. and Mohanty, Blue schist facies metamorphism in the ophiolitic melange of late Proterozoic Delhi Fold Belt, Rajasthan, India. *Precambrian Res.*, 1988, **42**, 97–105.
15. Gopalan, K., Macdougall, J. D., Roy, A. B. and Murali, A. V., Sm–Nd evidence for 3.3 Ga old rocks in Rajasthan, northwestern India. *Precambrian Res.*, 1990, **48**, 287–297.
16. Sharma, R. S., Geotectonic evolution of the Aravalli Mountain Belt. In Precambrian of the Aravalli Mountain Rajasthan, India (ed. Roy, A. B.), *Mem. Geol. Soc. India*, 1988, **7**, 33–75.
17. Tryggvason, A. *et al.*, Crustal architecture of the southern Uralides from true amplitude processing of the Urals Seismic Experiment and Integrated Studies (URSEIS) vibroseis profile. *Tectonics*, 2001, **20**, 1040–1052.
18. Cassell, B. R., A method for calculating synthetic seismograms in laterally varying media. *Geophys. J. R. Astron. Soc.*, 1982, **69**, 339–354.
19. Mishra, D. C., Singh, B., Tiwari, V. M., Gupta, S. B. and Rao, M. B. S. V., Two cases of continental collisions and related tectonics during the Proterozoic period in India – insights from the gravity modelling constrained by seismic and magnetotelluric studies. *Precambrian Res.*, 2000, **99**, 149–169.
20. Christensen, N. I. and Mooney, W. D., Seismic velocity structure and composition of the continental crust – a global view. *J. Geophys. Res. B*, 1995, **100**, 9761–9788.
21. Pareek, H. S., Petrochemistry and petrogenesis of the Malani Igneous suite, India: Summary. *Geol. Soc. Am. Bull.*, **92**, 67–70; **2**, 206–273.
22. Kochar, N. and Dhar, S., The association of hypersolvus and subsolvus granites: a study of Malani Igneous suite, India. *J. Geol. Soc. India*, **42**, 449–467.
23. Rudnick, R. L. and Fountain, D. M., Nature and composition of the continental crust: a lower crustal perspective. *Rev. Geophys.*, 1995, **33**, 267–307.
24. Mooney, W. D. and Brocher, T. M., Coincident seismic reflection/refraction studies of continental lithosphere. *Rev. Geophys.*, 1987, **25**, 723–742.
25. Meissner, R., Rupture, creep, lamellae and crocodiles happenings in continental crust. *Terranova*, 1989, **1**, 17–28.
26. Meissner, R. and Brown, L., Seismic reflection from earth's crust: comparative studies of tectonic patterns. *Geophys. J. Int.*, 1991, **105**, 1–2.
27. Page, R. A., Plafker, G., Fuis, G. S., Nokleberg, W. J., Ambos, E. L., Mooney, W. D. and Campbell, D. L., Accretion and subduction tectonics in Chugach Mountains and copper river basin, Alaska: Initial results of TACT. *Geology*, 1986, **14**, 501–505.
28. Pandey, O. P. and Agarwal, P. K., Lithospheric mantle deformation beneath Indian cratons. *J. Geol.*, 1999, **107**, 683–692.

ACKNOWLEDGEMENTS. We thank Dr V. P. Dimri, Director, NGRI for constant encouragement and support. N.S. thanks CSIR for providing her with the Senior Research Fellowship. Discussions with Dr D. Sarkar, Dr P. Koteswara Rao, and Mrs K. Arora were useful. Figures were drawn by Sri B. P. S. Rana and Sri M. Shankariah.

Received 7 July 2003; revised accepted 2 December 2003

How effective is an extended Kalman filter for continuous yeast cultures affected by both inflow and measurement noise?

P. R. Patnaik

Institute of Microbial Technology, Sector 39-A, Chandigarh 160 036, India

The usefulness of the extended Kalman filter (EKF) as an on-line estimator of process variables is known for monotonic laboratory-scale fermentations. However, this has not been tested for oscillating cultures under non-ideal conditions representative of large bioreactors. So, in this study an EKF was applied for on-line filtering of simulated data of an oscillating continuous *Saccharomyces cerevisiae* culture with inflow and measurement noise. For better accuracy, the tuning of the EKF was updated over successive time slices such that deviations between the noise-affected and noise-free profiles were minimized during each interval. As shown by the concentrations of biomass and ethanol, noise

e-mail: pratap@imtech.res.in

disrupted periodicity in both, but oscillations close to the noise-free behaviour could be restored substantially by the EKF, thus suggesting its suitability for large non-ideal bioreactors with either monotonic or oscillating cultures.

MICROBIAL fermentations operated under production conditions are often subject to noise from two sources. One source is broadly the group of sensory devices for different concentrations, temperature, pH, etc. Each of these devices has a separate measurement noise, but for modelling and control they are conveniently grouped into one vector with, as explained later, one covariance matrix for the noises. The other source is the environment, and this noise usually enters through a feed stream. Obviously, therefore, batch fermentations are less susceptible to this kind of noise, but they are not totally insulated from environmental effects because of variables such as the stirrer speed and pH control by the dosing of an alkali or an acid. By comparison, continuous and fed-batch fermentations are directly under the influence of environmental variations. Nevertheless, many microbial cultivations are carried out in fed-batch or continuous mode because of economic, kinetic or physiological benefits[1,2]. For such fermentations, suitable filtering of noise-affected data is at the core of any control strategy.

The Kalman filter is possibly the most widely used technique to generate smooth, usable data from noise-affected measurements. Since its theory is well documented[3], only a brief outline is provided later for completeness. Applications of the Kalman filter to microbial processes have been mainly for on-line estimations of variables that are difficult or expensive to measure on-line and whose off-line estimations are time-consuming[4,5]. Although production-scale operations are more prone to disturbances, ironically most studies with Kalman filters have focused on more 'ideal' laboratory-scale fermentations. However, these applications span different organisms, objectives, modes of operation and the variables monitored. The early work has been reviewed by Lubbert and Simutis[6], who pointed out both the potential and some weaknesses of the Kalman filter; these were addressed by later studies.

The variety of later applications is apparent from the observation that Neeleman and co-workers[7,8] used the Kalman filter to estimate the respiratory coefficients and specific growth rates of insect cell cultures; Zhang and Su[9] employed it for intracellular protein estimations in plant cell cultures, and Simon[10] for the removal of noise from measurements of polychlorinated biphenol degradation by bacterial co-cultures. Holwill *et al.*[11] also focused on proteins, in particular the estimation of the fractional precipitation of alcohol dehydrogenase from clarified yeast homogenate. This variety is extended further by the work of Vargas *et al.*[12] with waste-water treatment; their interest was in developing a time-optimal control system.

While establishing the effectiveness of the Kalman filter for bioreactors, these studies (with the exception of Simon[10]) were also restricted to laboratory-scale vessels operated under reasonably ideal conditions that did not truly mimic real production-scale processes. An important feature that 'non-idealizes' large-scale fermentations is the incursion of disturbances or noise. They usually enter through an inlet stream, and therefore fed-batch and continuous cultivations are more susceptible to process noise than batch operations. However, both modes of operation can be affected by measurement noise, which is a feature of the measuring instrument rather than that of the biological process. While intelligently controlled noise can be beneficial to a microbial culture, uncontrolled or improperly controlled noise can be seriously detrimental. For instance, noise can drive a fermentation from monotonic to chaotic behaviour or initiate run-away behaviour from a stable performance[13–15].

Despite its prevalence and the fact that noise can have more damaging effects on cellular processes than on chemical processes, recognition and quantitative analyses of the effects of noise on microbial processes are of recent origin. Nevertheless, now it has been established that disturbances flowing from the environment can seriously undermine cell viability, reactor stability, productivity and selectivity[14–16]. These studies cover different microorganisms and different bioreactor operating modes, and thus indicate a general validity of the injurious effects of noise. However, they all cover fermentations whose noise-free profiles change monotonically with time. While such fermentations include many important examples, they do not include an important class of fermentation which displays steady oscillations in continuous cultures. Perhaps the best known of these are *Zymomomas mobilis* and *Saccharomyces cerevisiae*, both of which generate ethanol.

Oscillating cultures exhibit a rich variety of performances, depending on the operating conditions. Both regular and aperiodic oscillations are possible, chaotic behaviour and, under restricted conditions, monotonic profiles too are observed[17,18]. Thus, the role of noise in such cultures is both difficult and important to understand. It is difficult because noise can cause abrupt shifts in the nature of the fermentation, and it is important because, as previous work with monotonic fermentations has shown[14,15], harnessing the noise suitably can improve cell growth and productivity beyond that of a noise-free fermentation. So the effectiveness of a Kalman filter in modulating noise arising simultaneously from the feed stream and the measuring devices for an oscillating continuous fermentation has been investigated in this work.

Many biological processes exhibit temporal oscillations under certain conditions. Yeast glycolysis[19], the circadian rhythm[20] and the cell cycle[21] have been widely studied. Equally important but less intensively analysed are metabolic oscillations seen in continuous cultures

of some bacteria (*Z. mobilis*)[22] and yeasts (*S. cerevisiae*)[17].

From the perspective of both molecular genetics and industrial fermentation, *S. cerevisiae* is an important organism because (a) its biochemistry and physiology are well understood, (b) noninvasive methods of measurement are possible in continuous cultures, (c) it does not generate endotoxins, and (d) it is a source of many useful products. These aspects and the mechanisms and models for oscillatory behaviour have been reviewed recently[23].

Oscillations in continuous cultures of *S. cerevisiae* have been reported with different carbon sources (glucose, ethanol or acetaldehyde) and different operating conditions, notably the dilution rate and gas–liquid mass transfer rate. Both intracellular reactions and transport between the cells and the surrounding fluid contribute to oscillatory behaviour. Because their interactions are complex and not fully understood, models have either focused on the intracellular kinetics or combined transport equations with lumped kinetics. The latter approach is a more sensible compromise than ignoring transport effects, especially since the latter can be significant in large bioreactors[1].

This study and a preceding one[24] are based on a cybernetic model proposed by Jones and Kompala[18]. This model was preferred over others for a few reasons. First, the cybernetic approach formalizes the established evolutionary concept that microorganisms try to follow those metabolic pathways that are most favourable to their survival under the prevailing conditions. Secondly, it has greater physiological closeness, it is simpler than mechanistic models of comparable accuracy and it portrays most of the observed features. Thirdly, the Jones–Kompala model has been able to explain excursions between different patterns of oscillations as a consequence of changes in operating conditions or substrate composition or genetic manipulation, which many mechanistic models have struggled to elucidate.

Briefly, Jones and Kompala[18] identified three metabolic pathways by which *S. cerevisiae* may utilize the available carbon sources: glucose fermentation, ethanol oxidation and glucose oxidation. The choice of pathway depends on the culture conditions, primarily the dilution rate, dissolved oxygen concentration and gas–liquid mass transfer rate. Although ethanol is synthesized under anaerobic conditions in batch cultures, it can be formed in certain ranges of the dissolved oxygen concentration and mass transfer rate in oscillating continuous cultures[25]. In the cybernetic framework, the organism chooses the pathway that is most favourable to its survival. So, it may shift from one pathway to another under changing conditions, and Jones and Kompala[18] postulated that that dynamic competition among the pathways is the main cause of oscillations. Their model is summarized in the Appendix.

The Kalman filter is a set of mathematical equations that provide an efficient recursive solution of the least-squares type. The filter can provide estimations of past, present and future states of a system even when a precise model is not known. This feature is useful for microbial processes under non-ideal (realistic) conditions because models developed on laboratory data may become inapplicable or imprecise under the influence of disturbances and spatial gradients[1,26].

The basic Kalman filter addresses the problem of trying to estimate the state $\bar{x}$ of a discrete time-controlled process that is governed by the linear difference equation:

$$\bar{x}_k = \bar{A}\bar{x}_{k-1} + \bar{B}\bar{u}_k + \bar{w}_{k-1}, \tag{1}$$

with a measurement vector that follows:

$$\bar{z}_k = \bar{H}\bar{x}_k + \bar{v}_k. \tag{2}$$

In these and later equations, lower-case letters with overbars denote vectors, while similar capital letters denote matrices. Scalars do not have overbars. $(k-1)$ is the current instant of time and k is the point one time-step ahead. $\bar{w}_k$ and $\bar{v}_k$ represent the process noise and measurement noise respectively.

Previous studies[26–28] show that $\bar{w}_k$ and $\bar{v}_k$ may be represented as white noise with normal probability distributions:

$$p(w) \sim N(0, \bar{Q}), \tag{3}$$

$$p(v) \sim N(0, \bar{R}), \tag{4}$$

where $\bar{Q}$ and $\bar{R}$ are the respective covariance matrices.

Since eq. (1) applies to linear equations whereas many fermenation (and other biological) processes follow nonlinear models, the extended Kalman filter (EKF) was developed. It applies to any nonlinear difference equation of the form:

$$\bar{x}_k = f(\bar{x}_{k-1}, \bar{u}_k, \bar{w}_{k-1}), \tag{5}$$

$$\bar{z}_k = \bar{h}(\bar{x}_k, \bar{v}_k). \tag{6}$$

In principle, the EKF solves the problem of determining the current estimates of a set of variables by expressing them as linear functions centred around the partial derivatives of the process and measurement functions evaluated at the (known) previous instant of time. The detailed theory and equations are given in the literature[3,29]. Note that both eqs (1) and (2) and eqs (5) and (6), in pairs, are in discrete form, whereas most biological processes are described by continuous models. This is not an impediment because, in practice, data are sampled at discrete points in time. Since the EKF allows any arbitrary variation in the sampling interval, this may be varied according to the nature of the process. For instance, the interval may be made inversely proportional to the current con-

centration gradient, thus generating closely spaced data when the variations are steep and more widely separated points during mild variations[14].

Earlier studies[6,14,16,28] have suggested that the feed stream is a major carrier of noise in continuous and fed-batch fermentations, and white noise is the principal component of the observed fluctuations. So, to generate data simulating a noise-influenced oscillating culture, the equations in the Appendix were solved with the parameter values used by Jones and Kompala[18] (see Table 1) and white noise specified by $\bar{Q}$ and $\bar{R}$. In an experimental application, the measurement covariance $\bar{R}$ is usually measured prior to the operation of the filter since it relates to the filter and not the process. The process noise covariance $\bar{Q}$ is more difficult to determine, since typically we do not have the ability to observe the process we are estimating. So, based on previous studies[5,10,29,30], $\bar{Q}$ was set initially to $\bar{Q}_d = ([0.0001 \ldots 0.0001]^T)$ and $\bar{R}$ to $0.003\bar{I}$, where $\bar{I}$ is the identity matrix and $\bar{Q}_d$ is a diagonal matrix. Now, the model of Jones and Kompala[18] has eight concentrations whose rates of change are expressed by eqs (A6)–(A11) in the Appendix. So $\bar{R}$ is an (8*8) matrix. Since the glucose and oxygen feed-streams are the only inflows to the bioreactor, environmental noise was considered to be present in these two flow rates, thus making $\bar{Q}$ a (2*2) matrix. Both $\bar{Q}$ and $\bar{R}$ get updated recursively as shown in Figure 1.

Apart from its applicability to nonlinear process, an important distinction between the EKF and the basic discrete Kalman filter is that in the former case the Jacobian $\bar{H}_k$, in the equation for the Kalman gain $\bar{K}_k$ also gets updated with each iteration, thereby speeding up convergence and improving the accuracy of estimations. For better filtering effectiveness, the duration of fermentation was divided into 1 h intervals and the tuning of the EKF was updated progressively over successive intervals.

Table 1. Values of parameters[18]

Parameter	Units	Value
α	h^{-1}	1.0
α^*	$g\ h^{-1}$	0.1
β	h^{-1}	0.2
γ_1	$g\ g^{-1}$	6.0
γ_2	$g\ g^{-1}$	6.0
γ_3	$g\ g^{-1}$	0.3
$\mu_{1,max}$	h^{-1}	0.44
$\mu_{2,max}$	h^{-1}	0.32
$\mu_{3,max}$	h^{-1}	0.31
φ_1	$g\ g^{-1}$	0.27
φ_2	$g\ g^{-1}$	1.067
φ_3	$g\ g^{-1}$	2.087
φ_4	$g\ g^{-1}$	0.95
D	h^{-1}	0.16
G_0	$g\ l^{-1}$	28.0
$k_L a$	h^{-1}	1200.0
K_1	$g\ l^{-1}$	0.1
K_2	$g\ l^{-1}$	0.02
K_3	$g\ l^{-1}$	0.001
K_{O2}	$mg\ l^{-1}$	0.0001
K_{O3}	$mg\ l^{-1}$	0.0001
O^*	$mg\ l^{-1}$	7.5
Y_1	$g\ g^{-1}$	0.16
Y_2	$g\ g^{-1}$	0.74
Y_3	$g\ g^{-1}$	0.50

The effects of noise and filtering are portrayed here through the concentrations of two key variables: biomass and ethanol. These variables were chosen because: (a) their magnitudes and amplitudes of oscillation are much larger than those of the other concentrations[18] and (b) cultivation processes generally try to maximize the productivity of these two variables.

S. cerevisiae also synthesizes and consumes internal storage carbohydrates, principally glycogen, trehalose and mannan[25,31]. In the work of Jones and Kompala[18], the carbohydrate concentration oscillated at an amplitude comparable to that of ethanol and at the same frequency. So the effect of noise on carbohydrate concentration was similar to that on ethanol, and hence only the ethanol profiles have been displayed. However, both are discussed in the context of the effect of noise on the metabolism and control of the process.

Figure 1 shows that there are two stages in the functioning of an EKF: initial a priori estimates are corrected by a posteriori estimates. In the present application, the a priori values pertained to pre-filtered noisy data and the corrected values were the result of filtering. The duration of fermentation represented in the study by Jones and Kompala[18], 50 to 76 h from the start, was divided into 26 one-hour intervals. During each interval the values of the covariances $\bar{Q}$ and $\bar{R}$ were determined such that the average deviation between the noise-affected and noise-free profiles was minimized. Since the concentrations at the end of each time slice depend on the (optimum) values of the previous interval, the feedback loop in Figure 1 is completed. This Markovian strategy not only conforms to the EKF theory but also has biological support, since its effectiveness has been demonstrated in other applications[14,16,32].

The extent of distortion of the concentration profiles is evident from Figure 2 *a* and *b*. Constancy of both amplitude and cycle time are disrupted by noise, and smooth oscillations degenerate to random multi-modal fluctuations. It is also seen that the biomass is more severely affected than ethanol concentration. These differences have important implications metabolically and for reactor control. Studies by different groups[25,31,33] have shown that the synthesis of biomass, ethanol and carbohydrates varies at different rates across the S, G1, G2 and M phases, but no consistent pattern seems to have been determined. Based on the work by Kuenzi and Fiechter[34], Jones and Kompala[18] have stated that *S. cerevisiae* stores carbohydrates when the medium is deficient in glucose and ethanol, and consumes the surplus carbohydrates when either

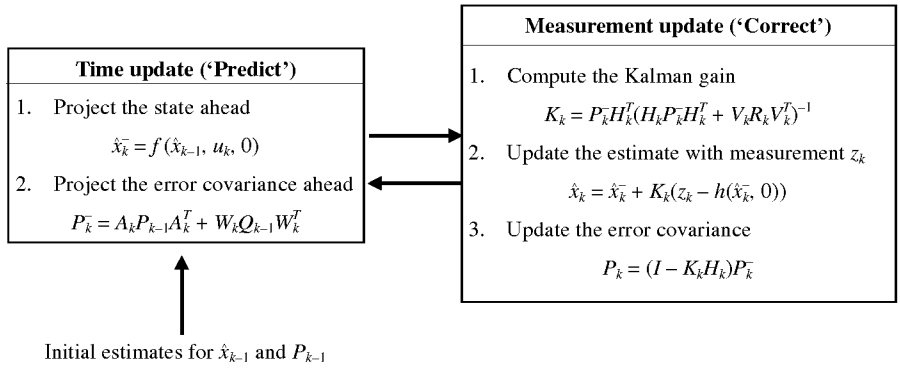

Figure 1. Computation procedure of the extended Kalman filter.

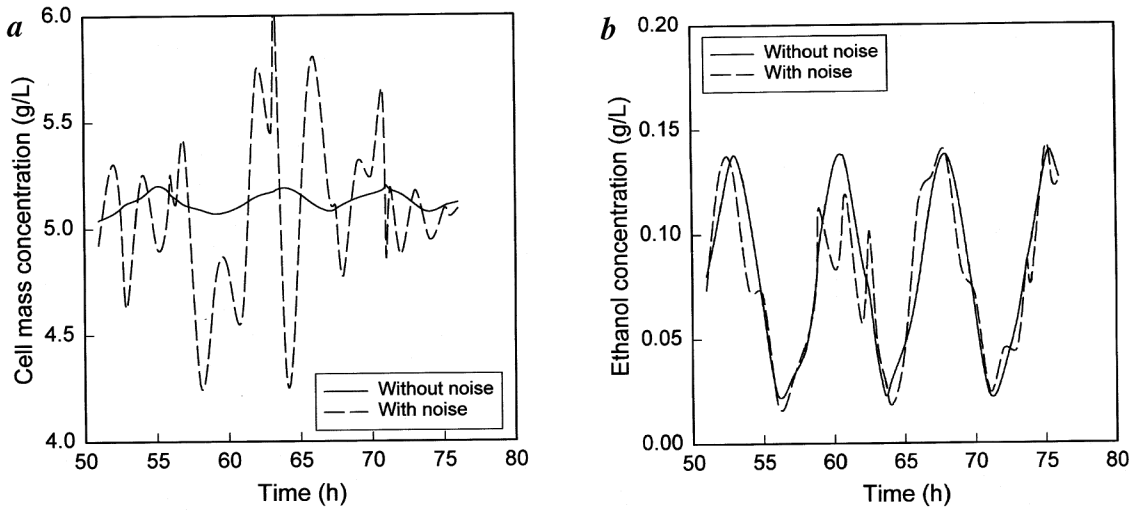

Figure 2. Time-domain profiles of (*a*) cell mass concentration and (*b*) ethanol concentration without noise and with noise.

substrate is present in large concentrations. Experiments by Duboc *et al.*[31] indicated a somewhat different inference. Biomass synthesis rate was high and ethanol synthesis low in all phases except the S phase, but carbohydrate synthesis was fast only in the G1 phase. Moreover, carbohydrates were consumed only in the G1 phase, whereas according to Jones and Kompala[18], there should have been some consumption in all phases, although at different rates.

Differences have also been observed between the synthesis and utilization rates of different carbohydrates. However, because of the lack of sufficient kinetic data for the individual constituents, it has not been possible to propose models for each of them and study the effect of noise. Nevertheless, it may be useful to note that glycogen, which constitutes the largest fraction of the carbohydrates, follows the same pattern as the biomass, whereas other carbohydrates and proteins do not[31]. The complexity of such variations, even without noise, requires sensory and control methods that are versatile, sensitive, fast, robust and 'intelligent'. Off-line measurements can be too slow to provide data for timely updating and corrective action, while on-line methods may be too expensive for production applications[4]. In such situations, knowledge-based methods, utilizing expert systems, neural networks and genetic algorithms meet these requirements much better than off-line sensors and PID-based control methods[35,36].

The effect of introducing an EKF is revealed in Figure 3 *a* and *b*. Since the actual profiles of the concentrations have been shown in Figure 2 *a* and *b*, the percentage deviations are now more informative. Corresponding to the smaller extent of distortion by noise for ethanol concentration than for biomass, the extent of improvement produced by the EKF is also larger for ethanol. What is visually apparent from Figure 3 *a* and *b* may be quantitatively characterized by defining a filtering index (FI):

$$\mathrm{FI} = \sum_{i=1}^{M} \frac{(\%\ \text{deviation without filter})_i - (\%\ \text{deviation with filter})_i}{(\%\ \text{deviation without filter})_i}, \quad (7)$$

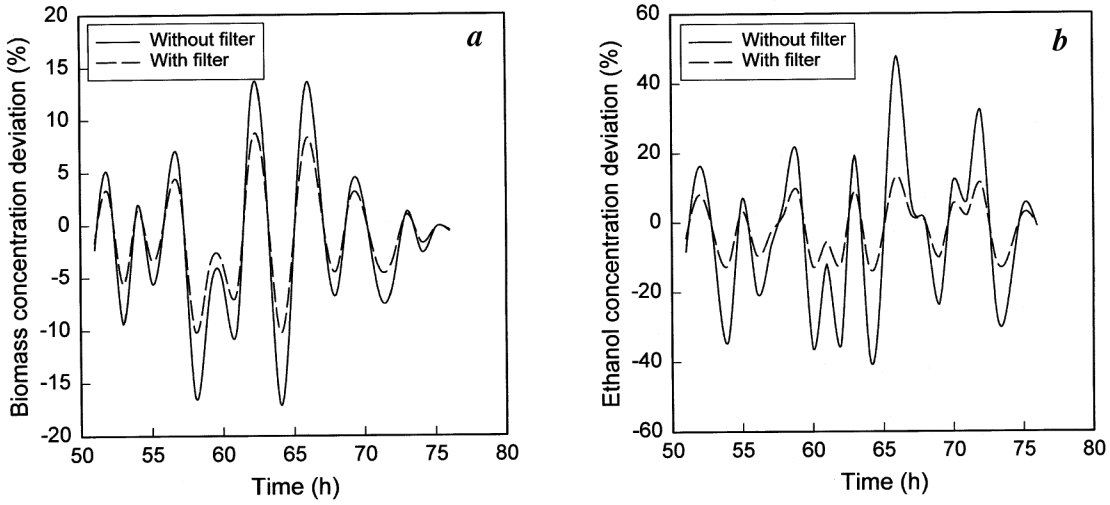

Figure 3. Variation in the deviation of noise-affected concentration of (*a*) biomass and (*b*) ethanol (from the noise-free concentration) without and with an extended Kalman filter.

Table 2. Pearson's product moment correlation coefficients

	Biomass (unfiltered)	Ethanol (unfiltered)	Biomass (filtered)	Ethanol (filtered)
Biomass (unfiltered)	1.0	0.166	0.999	0.122
Ethanol (unfiltered)	–	1.0	0.162	0.984
Biomass (filtered)	–	–	1.0	0.122

where M is the number of datapoints. In this study, M was 26. The calculated values of FI were 0.563 for ethanol and 0.338 for biomass. The values may be improved by employing (a) shorter time intervals, (b) two or more filters in series or (c) a combination of an EKF and a knowledge-based device[5,30]. However, each of these methods increases the computation effort, and consequently the response time at each stage. So the choice of the best filtering technique rests on a balance between the improvement achieved and the effort required[14,37]; this aspect is under investigation. At the theoretical limit of a perfect filter, the noise-free performance should be fully recovered and hence FI $\rightarrow$ 1.

To determine whether the deviations without and with an EKF were correlated, Pearson's product moment correlations and scatter diagrams were computed as described by Fisher[38]. The correlation between two noise-influenced variables X and Y is defined as:

$$\rho_{x,y} = \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)\,\mathrm{var}(Y)}, \tag{8}$$

where var stands for variance and cov for covariance. The correlation coefficients (Table 2) show that, as expected, noise-induced deviations in the concentrations of biomass and ethanol are not correlated (small coefficients), whereas each is strongly correlated with itself.

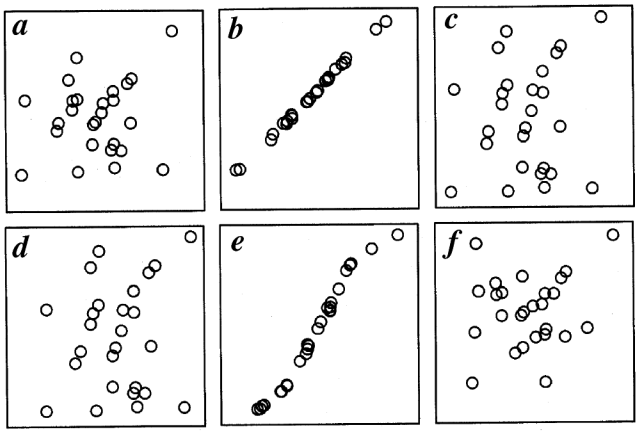

Figure 4. Scatter diagrams portraying the relation between the deviations of pairs of variables at different times. Abscissae and ordinates correspond to the following variables: *a*, XU, EU; *b*, XU, XF; *c*, XU, EF; *d*, XF, EF; *e*, EU, EF; *f*, XF, EU; XU, Unfiltered biomass; XF, Filtered biomass; EU, Unfiltered ethanol; EF, Filtered ethanol.

Whereas correlation coefficients present an overall picture, detailed variations in the correlation between the concentrations at each of the 26 sampled points are portrayed in the scatter diagrams (Figure 4). Each set of points relates the deviations caused by noise (without or with filtering) for one viable with another. For instance, Figure 4 *a* relates unfiltered biomass concentration with unfiltered ethanol concentration. As expected, the points are randomly scattered here and in Figure 4 *c*, *d*, and *f*, all relating biomass and ethanol concentrations under different conditions. These random distributions translate to the small correlation coefficients in Table 2. On the contrary, the deviations for the same variable without and with an EKF (Figure 4 *b* and *e*) are distributed linearly, indicating a strong correlation at all times, reflected in the large coefficients in Table 2.

The EKF has been shown in previous studies to be a useful on-line estimator of state variables for nonlinear fermentation processes. However, those processes were relatively 'ideal' in that there was minimum external noise (mainly because of the small scale of operation and elaborate controls) and the disturbances were mainly in the measurement procedure.

To evaluate the useful of an EKF in more difficult (and realistic) situations, this study considered an oscillating continuous fermentation with noise in both the feed stream and the measuring devices. Using an experimentally established cybernetic model for the continuous cultivation of *S. cerevisiae* in a noise-free bioreactor, computer-generated white noise was introduced in the feed stream and in the sensors for the performance variables by suitable covariance matrices, and the modified model was solved to generate data mimicking a large-scale, noise-affected fermentation.

Noise destroyed the smooth periodic variations in the concentrations of the biomass and the product ethanol. An EKF was employed to filter out the noise and recover the deterministic performance. Based on previous work, the EKF was tuned stepwise over successive slices of time covering the duration of fermentation. The noise-free, smooth oscillations were substantially recovered with an EKF, but there were differences in the extent of recovery for the two concentrations. The less affected concentration of ethanol was also recovered to a larger extent. Further improvements are possible, but they can slow down the response times.

Since this was a simulated experiment, randomness of the noise-induced displacements without and with an EKF was cross-checked by Pearson's correlation coefficients and scatter plots. Both showed, as expected, that displacements in the concentrations of ethanol and biomass were uncorrelated, whereas those for the same variable (without and with filtering) were strongly correlated. In summary, therefore, this study shows that an EKF can be an effective noise filter and state estimator in realistic conditions for complex fermentation processes.

Appendix

The cybernetic model of Jones and Kompala[18]

Depending on the prevailing conditions, *S. cerevisiae* may follow any one of three metabolic pathways. The rate of growth r_i along each pathway follows modified Monod kinetics, as given below.

Glucose fermentation

$$r_1 = \mu_1 e_1 \left(\frac{G}{K_1 + G} \right), \tag{A1}$$

Ethanol oxidation

$$r_2 = \mu_2 e_2 \left(\frac{E}{K_2 + E} \right) \left(\frac{O}{K_{O2} + O} \right), \tag{A2}$$

Glucose oxidation

$$r_3 = \mu_3 e_3 \left(\frac{G}{K_3 + G} \right) \left(\frac{O}{K_{O3} + O} \right). \tag{A3}$$

The pathways are not mutually exclusive and at a given instant, the organism may follow two or more pathways at different rates. Each pathway is controlled by a key enzyme e_i, with synthesis rate u_i and activity v_i, which follow:

$$u_i = \frac{r_i}{\sum_j r_j}, \tag{A4}$$

$$v_i = \frac{r_i}{\max_j r_j}. \tag{A5}$$

With eqs (A1)–(A5), the mass balances for a continuous flow bioreactor may be written as follows:

$$\frac{dX}{dt} = \left(\sum_i (r_i v_i) - D \right) X, \tag{A6}$$

$$\frac{dG}{dt} = (G_0 - G)D - \left(\frac{r_1 v_1}{Y_1} + \frac{r_3 v_3}{Y_3} \right) X - \varphi_4 \left(C \frac{dX}{dt} + X \frac{dC}{dt} \right), \tag{A7}$$

$$\frac{dE}{dt} = -DE + \left(\varphi_1 \frac{r_1 v_1}{Y_1} - \frac{r_2 v_2}{Y_2} \right) X, \tag{A8}$$

$$\frac{dO}{dt} = k_L a(O^* - O) - \left(\varphi_2 \frac{r_2 v_2}{Y_2} + \varphi_3 \frac{r_3 v_3}{Y_3} \right) X, \tag{A9}$$

$$\frac{de_i}{dt} = \alpha u_i \left(\frac{S_i}{K_i + S_i} \right) - \left(\sum_j (r_j v_j) + \beta \right) e_i + \alpha^*, \tag{A10}$$

$$\frac{dC}{dt} = \gamma_3 r_3 v_3 - (\gamma_1 r_1 v_1 + \gamma_2 r_2 v_2) C - \sum_j (r_j v_j) C. \tag{A11}$$

Inclusion of the term α^* in the enzyme synthesis equations (A10) is based on Turner and Ramkrishna[39], who have shown its importance in predicting the induction of

enzymes that have been repressed for long durations. The specific growth rates thus also include α^* in the model:

$$\mu_i = \mu_{i,\max}\left(\frac{\mu_{i,\max} + \beta}{\alpha + \alpha^*}\right). \quad \text{(A12)}$$

Equation (A11) expresses the rate of change of internal storage carbohydrates that are an integral part of the metabolism[25,31].

The φ_i are stoichiometric coefficients for different substrates S_i, and γ_i are similar coefficients for carbohydrate synthesis and consumption by the cells. Jones and Kompala[18] may be consulted for a full discussion of the model. A point not clarified there is the identification of S_1, S_2 and S_3. Reference to eqs (A1)–(A3) shows that $S_1 = G$, $S_2 = E$ and $S_3 = G$. This identification is needed to solve the model. The values of the parameters are listed in Table 1.

Nomenclature

C = Intracellular storage carbohydrate concentration (g l^{-1})
D = Dilution rate (h^{-1})
e_i = Key enzyme concentration for ith pathway (g g^{-1} biomass)
E = Ethanol concentration in the bioreactor (g l^{-1})
G = Glucose concentration in the bioreactor (g l^{-1})
G_0 = Glucose concentration in the feed stream (g l^{-1})
$k_L a$ = Oxygen mass transfer coefficient (h^{-1})
K_i = Michaelis constant for ith pathway (g l^{-1})
K_{O2}, K_{O3} = Oxidative pathway oxygen saturation constants (mg l^{-1})
O = Dissolved oxygen concentration in the bioreactor (mg l^{-1})
O^* = Dissolved oxygen solubility limit (mg l^{-1})
r_i = Biomass growth rate on ith pathway (h^{-1})
S_i = Carbon substrate concentration for ith pathway (g l^{-1})
t = Elapsed time (h)
u_i = Cybernetic variable controlling key enzyme synthesis for ith pathway (–)
v_i = Cybernetic variable controlling key enzyme activity for ith pathway (–)
X = Biomass concentration in the bioreactor (g l^{-1})
Y_i = Yield coefficient for ith pathway (g biomass g^{-1} substrate)

α = Specific enzyme synthesis rate (h^{-1})
α^* = Constitutive enzyme synthesis rate (g h^{-1})
β = Specific enzyme degradation rate (h^{-1})
φ_i = Stoichiometric coefficient for ith carbon substrate (–)
γ_i = Stoichiometric coefficients for storage carbohydrate synthesis and degradation (–)
μ_i = Specific growth rate of biomass on ith substrate (h^{-1})
$\mu_{i,\max}$ = Maximum specific growth rate on ith substrate (h^{-1})

1. Liden, G., Understanding the bioreactor. *Bioproc. Biosyst. Eng.*, 2002, **24**, 273–279.
2. Shuler, M. L. and Kargi, F., *Bioprocess Engineering. Basic Concepts*, Prentice-Hall, New Jersey, 2002.
3. Grewal, M. S. and Andrews, A. P., *Kalman Filtering Theory and Practice*, Prentice Hall, New Jersey, 1993.
4. Locher, G., Sonnleiter, B. and Fiechter, A., On-line measurement in biotechnology: techniques. *J. Biotechnol.*, 1992, **25**, 23–53.
5. Soroush, M., State and parameter estimations and their applications in process control. *Comput. Chem. Eng.*, 1998, **23**, 229–245.
6. Lubbert, A. and Simutis, R., Using measurement data in bioprocess modelling and control. *Trends Biotechnol.*, 1994, **12**, 304–311.
7. Neeleman, R. and van Boxtel, A. J. B., Estimation of specific growth rate from cell density measurements. *Bioproc. Biosyst. Eng.*, 2001, **24**, 179–185.
8. Neeleman, R., van den End, E. J. and van Boxtel, A. J. B., Estimation of respiratory coefficient in batch cell cultivation. *J. Biotechnol.*, 2000, **80**, 85–95.
9. Zhang, J. N. and Su, W. W., Estimation of intra-cellular phosphate content in plant cell cultures using extended Kalman filter. *J. Biosci. Bioeng.*, 2002, **94**, 8–14.
10. Simon L., Observing biomass concentration profiles in a fixed-bed reactor. Paper No. 318j, AIChE Annual Meeting, 6 November 2002.
11. Holwill, I., Chard, S., Flanagan, M. T. and Hoare, M., A Kalman filter algorithm and monitoring apparatus for at-line control of fractional protein precipitation – systems control. *Biotechnol. Bioeng.*, 1997, **53**, 58–70.
12. Vargas, A., Soto, G., Moreno, J. and Buitron, G., Observer-based time-optimal control of an aerobic SBR for chemical and petrochemical waste treatment. *Water Sci. Technol.*, 2000, **42**, 163–170.
13. Patnaik, P. R., Fractal characterization of the effect of noise on biological oscillations: the biosynthesis of ethanol. *Biotechnol. Tech.*, 1994, **8**, 419–424.
14. Patnaik, P. R., Improvement of the microbial production of streptokinase by controlled filtering of inflow noise. *Process Biochem.*, 1999, **35**, 309–315.
15. Patnaik, P. R., Can imperfections help to improve bioreactor performance? *Trends Biotechnol.*, 2002, **20**, 135–137.
16. Patnaik, P. R., Spectral analysis of the effect of inflow noise on a fed-batch fermentation for recombinant β-galactosidase. *Bioprocess Eng.*, 1997, **17**, 93–97.
17. Chen, C.-I. and McDonald, K. A., Oscillatory behavior of *Saccharomyces cerevisiae* in continuous culture, I and II. *Biotechnol. Prog.*, 1990, **36**, 28–38.
18. Jones, K. D. and Kompala, D. S., Cybernetic model of growth dynamics of *Saccharomyces cerevisiae* in batch and continuous cultures. *J. Biotechnol.*, 1999, **71**, 105–131.
19. Richard, P., Bakker, B. M., Teusink, B., van Dam, K. and Westerhoff, H. V., Acetaldehyde mediates the synchronization of sustained glycolytic oscillations in populations of yeast cells. *Eur. J. Biochem.*, 1996, **235**, 238–241.
20. Turek, F. W., Circadian rhythms. *Horm. Res.*, 1998, **49**, 109–113.
21. Mori, T. and Johnson, C. H., Circadian control of cell division in unicellular organisms. *Prog. Cell Cycle Res.*, 2000, **4**, 185–192.

22. Bruce, L. J., Axford, D. B., Ciszek, B. and Daugulis, A. J., Extractive fermentation by *Zymomonas mobilis* and the control of oscillatory behavior. *Biotechnol. Lett.*, 1991, **13**, 291–296.
23. Patnaik, P. R., Oscillatory metabolism of *Saccharomyces cerevisiae*: mechanisms and models. *Biotechnol. Adv.*, 2003, **21**, 183–192.
24. Patnaik, P. R., On the performances of noise filters in the restoration of oscillatory behavior in continuous yeast cultures. *Biotechnol. Lett.*, 2003, **25**, 681–985.
25. Satroutdinov, A. D., Kuriyama, H. and Kobayashi, H., Oscillatory metabolism of *Saccharomyces cerevisiae* in continuous cultures. *FEMS Microbiol. Lett.*, 1992, **98**, 261–268.
26. Gillard, F. and Tragardh, C., Modeling the performance of industrial bioreactors with a dynamic micro-environmental approach: a critical review. *Chem. Eng. Technol.*, 1999, **22**, 187–195.
27. Montague, G. A. and Morris, A. J., Neural network contributions in biotechnology. *Trends Biotechnol.*, 1994, **12**, 312–324.
28. Rohner, M. and Meyer, H.-P., Application of modeling for bioprocess design and control in industrial production. *Bioproc. Eng.*, 1995, **13**, 69–78.
29. Stephanopoulos, G. and Park, S., Bioreactor state estimation. In *Biotechnology Measuring, Modelling and Control* (ed. Schugerl, K.), VCH, Weinheim, 1992, vol. 4, pp. 225–249.
30. Zorzetto, L. F. M. and Wilson, J. A., Monitoring bioprocesses using hybrid models and an extended Kalman filter. *Comput. Chem. Eng.*, 1996, **20**, S689–S694.
31. Duboc Ph., Marison, I. and von Stochar, U., Physiology of *Saccharomyces cerevisiae* during cell cycle oscillations. *J. Biotechnol.*, 1996, **51**, 57–72.
32. Glassey, J., Montague, G. A., Ward, A. C. and Kara, B. V., Artificial neural network based experimental design procedures for enhancing fermentation development. *Biotechnol. Bioeng.*, 1994, **44**, 399–405.
33. Martegani, E., Porro, D., Ranzi, B. M. and Alberghina, L., Involvement of a cell size control mechanism in the induction and maintenance of oscillations in continuous cultures of budding yeast. *Biotechnol. Bioeng.*, 1990, **36**, 453–459.
34. Kuenzi, M. T. and Fiechter, A., Changes in carbohydrate composition and trehalose activity during the budding cycle of *Saccharomyces cerevisiae*. *Arch. Mikrobiol.*, 1969, **64**, 396–407.
35. Berkholz, R., Rohlig, D. and Guthke, R., Data and knowledge based experimental design for fermentation process optimization. *Enzyme Microb. Technol.*, 2000, **27**, 784–788.
36. Shioya, S., Shimizu, K. and Yoshida, T., Knowledge-based design and operation of bioprocess systems. *J. Biosci. Bioeng.*, 1999, **87**, 261–266.
37. Patnaik, P. R., Enhancement of protein activity in a recombinant fermentation by optimization of fluid dispersion and initial plasmid copy number distribution. *Biochem. Eng. J.*, 2001, **9**, 111–118.
38. Fisher, L., Probability and statistics. In *Handbook of Applied Mathematics* (ed. Pearson, C. E.), Van Nostrand, New York, 1990, chapter 21.
39. Turner, B. G. and Ramkrishna, D., Revised enzyme synthesis rate expression in cybernetic models of bacterial growth. *Biotechnol. Bioeng.*, 1988, **31**, 41–43.

Received 14 August 2003, accepted 5 November 2003

Cloning, characterization and expression of *AmphiCypA*, a homologue of eukaryotic cyclophilin A gene from amphioxus *Branchiostoma belcheri tsingtauense*

L. Wang[1], S. Zhang[1,*], Z. Liu[1], H. Li[1] and A. Xu[2]

[1]Department of Marine Biology, Ocean University of China, Qingdao 266 003, P.R. China
[2]College of Life Science, Zhongshan University, Guangzhou 510 275, P.R. China

This study reports the cloning and characterization of a full-length amphioxus cyclophilin cDNA. The cDNA consisted of 1358 bp with a 495-bp open reading frame (ORF) corresponding to a deduced protein of 164 amino acids, with a calculated molecular mass of 17.3 kDa. It possesses the signature sequence of Cyclophilins (Cyps) in the PROSITE library, and all the conserved 13 amino acids, including the crucial tryptophan (W) residue (position 121) that are predicted to be involved in PPIase activity and CsA binding. In addition, it has neither N-terminal extension sequence nor C-terminal extension sequence. These indicate that the cDNA encodes a homologue of eukaryotic cyclophilin A, designated as *AmphiCypA*. Phylogenetic analysis shows that AmphiCypA is intermediate between sea urchin Cyp and zebrafish Cyp, clustering together with zebrafish Cyp. This agrees with the notion that amphioxus represents a basal lineage of chordates in phylogeny. Northern blot analysis revealed that *AmphiCypA* is expressed in all the tissues examined, but its expression is apparently elevated in fully-grown ovaries. It is suggested that elevated expression of *AmphiCypA* in fully-grown ovaries could be due to a role for this protein in oogenesis or it may be involved in early development.

CYCLOPHILINS (Cyps) are a family of proteins that bind to the immunosuppressive agent, cyclosporin A (CsA), via a central highly conserved CsA-binding domain[1–3]. Biochemical studies have shown that Cyps have peptidyl-prolyl *cis–trans* isomerase (PPIase: EC 5.2.1.8) activity[4,5]. Their biological significance is manifested by the catalysis of protein folding via peptide bond rotation on the amino side of proline residues[6,7], the action as a chaperone for protein trafficking[8], the nucleolytic degradation of the genome[9] as well as the involvement in stress response[10,11].

The first identified cyclophilin was human CypA, an 18-kDa soluble cytoplasmic protein[12]. To date, divergent types of Cyps have been identified on the basis of their size and target location. For example, in addition to CypA, at least four other types of Cyps have been characterized

*For correspondence. (e-mail: zscshab@public.qd.sd.cn)