

# Protein folding and the robustness of cells

G.J. Shawn Tan, Ferran D. Revilla, Klaus-Peter Zauner\*

*School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, United Kingdom*

Received 28 February 2005; received in revised form 8 July 2006; accepted 15 July 2006

## Abstract

The intricate intracellular infrastructure of all known life forms is based on proteins. The folded shape of a protein determines both the protein's function and the set of molecules it will bind to. This tight coupling between a protein's function and its interconnections in the molecular interaction network has consequences for the molecular course of evolution. It is also counter to human engineering approaches. Here we report on a simulation study investigating the impact of random errors in an abstract metabolic network of 500 enzymes. Tight coupling between function and interconnectivity of nodes is compared to the case where these two properties are independent. Our results show that the model system under consideration is more robust if function and interconnection are intertwined. These findings are discussed in the context of nanosystems engineering.

© 2006 Elsevier Ireland Ltd. All rights reserved.

**Keywords:** Self-isolation; Organic devices; Complex networks; Nanoengineering; Evolvability; Modularity

## 1. Nature's nanoengineering

Nature's astounding integration density has its roots in an enormous number of macromolecular structures that implement a wide variety of highly specific functions. Most prominent among them are proteins. These ubiquitous functional nanocomponents are macromolecules consisting of several thousand atoms. Biomimicry in the field of materials sciences is increasingly exploring the molecular and nanolevel detail that yields the remarkable properties of animate matter. A technology that would enable the deliberate design and fabrication of similar materials and systems would be highly desirable for a broad variety of applications (Lehn, 2002). Established engineering processes, however, are not well suited to the development of

organic devices (Ball, 2002; Luo, 2003). In the context of such a technology biological means of achieving robust systems are of interest. In the following we will explore a potential role of protein folding for the robustness of networks of interacting proteins.

### 1.1. Protein folding

Proteins are linear chains of typically a few hundred building blocks taken since two thousand million years from the same set of about 20 amino acids. The linear chain is assembled according to an edited molecular copy of the coding region of a gene on the DNA. In principle the amino acids can be combined in arbitrary order. This opens up a vast space of possible macromolecules that can be assembled from the amino acid building blocks. During and following the assembly, the protein will spontaneously curl up under the electrostatic interaction of its atoms into a defined but agile three-dimensional form.

In proteins function follows form. Folding of the amino acid chain into a spacial structure is essential to the capabilities of a protein. The resulting spatial

\* Corresponding author. Tel.: +44 23 8059 3089; fax: +44 23 8059 9179.

E-mail addresses: [nwahs81@hotmail.com](mailto:nwahs81@hotmail.com) (G.J. Shawn Tan), [fdr04r@ecs.soton.ac.uk](mailto:fdr04r@ecs.soton.ac.uk) (F.D. Revilla), [kpz@ecs.soton.ac.uk](mailto:kpz@ecs.soton.ac.uk) (K.-P. Zauner).

structure typically fits in a sphere of 4–8 nm diameter and is largely determined by the amino acid sequence. Amino acids differ in geometry, electrostatic characteristics, and chemical properties. Not all amino acid chains readily fold (Sosnick et al., 2002), but many randomly created sequences do (Davidson and Sauer, 1994). More often than not, however, random sequences aggregate and are insoluble in water (Prijambada et al., 1996). Proteins that occur in cells have sequences that allow them to fold under physiological conditions. Due to a considerable energy gap between the native and any competing fold, natural proteins fold in a cooperative all-or-none transition. This energy gap is the key characteristic that distinguishes a natural protein from a random amino acid sequence (Finkelstein and Galzitskaya, 2004). Electrostatic interactions among amino acid residues that approach closely in space facilitate correct folding (Dobson, 2003; Lindorff-Larsen et al., 2005). An important feature of the folded three-dimensional protein structure is the coupling of amino acids distant on the linear sequence but in close proximity in space. Through this coupling local changes in the protein are propagated through the protein's structure, a process common in the transduction of signals (Luque et al., 2002).

The final folded protein has a characteristic compact shape. A precise spatial placement and orientation of specific amino acid residues gives rise to the broad spectrum of functions exhibited by proteins. The highly selective and efficient catalytic properties of enzymes, for instance, rely on the precise positioning of atom groups in the catalytic centre. Furthermore, the shape of the protein is also crucial for its capability of interacting with other molecules (Fischer, 1894; Friedrich, 1984). The spatial conformation of the protein's amino acid chain thus determines both the function and the interactions of the protein.

### 1.2. Networks of proteins

In the cell proteins are embedded in molecular networks. If two proteins have mutual compatible docking surfaces, they can interact directly. Proteins can also interact through exchange of molecular signals. Enzymes, i.e. catalytically active proteins, can form metabolic networks where the reaction product of one enzyme serves as substrate for a following enzyme. In this paper we focus on metabolic networks. Metabolic networks, in contrast to signalling networks, are connected through precise steric fit of enzymes and substrates and therefore considered more difficult to evolve (Kirschner and Gerhart, 1998). Our interest here is the robustness of metabolic networks to faults in its components, i.e. in

the enzymes. The robustness of networks is frequently discussed from a solely topological perspective, for example, the connectedness of the network after random deletion of nodes is considered (Barabási and Oltvai, 2004). On the contrary, the approach taken here is concerned with dynamic properties of the metabolic network. To simulate the network dynamics we use an abstract artificial chemistry model (Dittrich et al., 2001) described in detail in Section 2.

### 1.3. Faulty proteins

Proteins are complex molecular machines the function of which typically hinges on positioning of atomic groups with Å-precision. This positioning is achieved through the self-organisation of folding. A modification of the amino acid sequence of a protein will affect the cooperative folding process, the spatial arrangement of atoms in the folded protein, and the conformational dynamics of the protein structure. Over the course of evolution the complex protein networks of cells are formed through diversification and specialisation of protein structures as a consequence of alterations in their amino acid sequences. Changes to the DNA will lead to inheritable modifications of protein sequences that form the basis of molecular evolution as well as genetic disease (Wang and Moulton, 2001). Errors can also be introduced at later stages in the production of proteins. An erroneous transcription of DNA into mRNA, for instance, will give rise to a “production run” of faulty proteins over the lifetime of the incorrect mRNA molecule. If a correct mRNA is incorrectly translated then individual proteins with faults are produced. The cell employs an elaborate molecular machinery for the quality control of proteins (Netzer and Hartl, 1998) and eliminates proteins that cannot fold correctly (Dobson, 2003).

Altering an amino acid in a protein can affect the protein in three basic ways:

- (1) It can derail the folding process, e.g. through steric obstruction or electrostatic repulsion or attraction.
- (2) It can destabilise the folded protein, e.g. by eliminating hydrogen bonds or salt bridges or by disrupting the non-polar core area.
- (3) It can eliminate a signal sequence pattern, e.g. a recognition signal for protein targeting or post-translational modification.

The first two mechanisms can impact ligand binding, catalytic activity, and regulation of the protein. Depending on the protein, the sequence position, and the amino acid substituted, the effect can range in severity

from neutral to catastrophic (Wang and Moulton, 2001). Investigations have shown that a single random amino acid substitution will result with 30–40% probability in a non-functional protein (Guo et al., 2004). On this basis it may be considered as probable that the majority of random amino acid changes can be absorbed on the molecular level (Conrad and Volkenstein, 1981) and result in near neutral sequence variation (Kimura and Ohta, 1974; Grishin, 2001). In the light of the discovery that the combination of substitutions with individually mild effects causes, with high probability, catastrophic loss of function (Axe, 2000), the question arises whether the context sensitivity conferred on the sequence by the folding process can impart an evolutionary advantage beyond mutation buffering (Conrad, 1979). Perhaps, a protein with a fault and thus possibly with a subtly altered function or regulation is more problematic than a protein that is altered drastically. The latter presumably can be recognised much more readily by the quality control mechanisms of the cell (Netzer and Hartl, 1998). Furthermore, a faulty protein with a significantly altered shape is likely to disengage from the molecular interaction network with no ill effect as long as functional copies of the protein are still available in the cell. In the following section we will describe an abstract model of a metabolic network developed to study such a scenario.

## 2. A metabolic network model

An artificial chemistry model suitable for studying the impact of random errors on enzymatic networks has been developed. A key requirement for the model is the ability to abstract the complex non-local interactions of protein folding process. The model assumes perfect mixing of the metabolites through diffusion, a reasonable assumption for a cell of micrometre scale and enzyme turn-over rates in the ms-range (Adam and Delbrück, 1968). Metabolites, here interchangeably also called substrates, are represented as molecule types and the count of molecules present in the cell for each such type. Similarly enzymes are represented as specific types and the count of enzymes present in the cell for each type of enzyme. In the following we will first describe the representation used for the chemical compounds, followed by the stochastic reaction rules that govern the dynamics of the model, and then discuss the abstraction of folding.

### 2.1. Substrate representation

Each substrate type is represented as a bit string composed of two parts. The first part describes the size of

the molecule and can be analogised to the carbon count of an organic molecule. In our virtual chemistry, however, only one type of atom exists and is accounted for. The remaining bits of the substrate's bit string represent the configuration of the atoms. In all experiments reported here the substrate types consist of 8 bit for the size and 10 bit for the molecular structure, i.e. all substrate types have representation 18 bit long. The number of bits used should open up a large space of potential molecules, however, the time required to generate connected networks (Section 2.3) increases with the space of potential molecules. The structure bits do not have any specific interpretation other than that molecules of identical size are different if and only if their structure bits differ. All reactions will take heed of conservation of mass and insure that the sum over the size (the bits interpreted as binary number) of the reactants is equal to the sum over the size of the products. Substrates with a size of zero are not allowed.

### 2.2. Enzyme abstraction

In this model enzyme types come in two categories. Either they process a single substrate or they catalyse binary reactions, accepting two substrates. An enzyme of the former kind is specified by a bit string twice as long as the bit string of a substrate type. This bit string can be decomposed into two parts, a recognition part and an operation part. The recognition part is identical to the bit string of the substrate type on which the enzyme operates. The operation part is a bit string of the same length and the same format as a substrate type but is interpreted in a different way. The recognition and operation parts of the enzyme type implicitly define the product of the reaction catalyzed by the enzyme through the following rule. If the size of the substrate is larger than the size specified in the operation bits of the enzyme, the reaction will have two products. One product will have the size specified in the size bits of the enzyme's operation, the other product will have the remaining mass, i.e. the difference between the size specified in the recognition part of the enzyme and the size specified in the operation part. The configuration part of both products are set to the same bit string: the result of a bitwise exclusive-or (XOR) between the configuration bits of the recognition part and the configuration bits of the operation part. If the substrate size is less than the size specified in the enzyme's operation part, only one product of the same size as the substrate is formed. The configuration of this product is set to the XOR of the substrate configuration with the configuration bits of the enzyme's operation part. The XOR operation is chosen because it is symmetric

and hence does not constrain the possible structure of products of long reaction chains.

Enzymes that catalyse binary reactions are specified by bit strings three times as long as a substrate type. This specification corresponds to the one described above extended by a second substrate recognition part. The sizes of the products are determined by comparing the sum of the sizes of the substrates with the enzyme's operation size. If the operation size is larger than this sum, one product is formed. Otherwise the reaction has two products, one of which has the size specified by the enzyme's operation part, the other receives the remaining mass. To make this rule account for substrate and product sizes correctly, enzyme types in this category have to fulfil the condition  $s_{op} > 1 + s_{s1} + s_{s2} - 2^k$ , where  $k$  is the number of bits used to express size (here 8), and  $s_{op}$ ,  $s_{s1}$  and  $s_{s2}$  are the size of the enzymes' operation, first substrate and second substrate, respectively. The configuration bits of the products are set to the XOR of the configuration bits of the two substrates if only one product is formed. If two metabolites are produced, one product has configuration bits formed with XOR from the operation bits and the first substrate while the other product is formed with the operation bits and the second substrate.

With this enzyme specification, anabolic reactions that combine two substrates into one product as well as catabolic reactions that split substrates are possible. For simplicity, in the simulations reported here all enzymes are assumed to have identical rate constants. Defects in enzymes (cf. Section 1.3) are modelled by inverting bits in the bit string that describes the enzyme type.

A fixed set of substrates is provided to the metabolic network. This set comprises all substrates recognised by the enzymes of the network minus all substrates produced by the enzymes of the network. Reactions commence by selecting a random enzyme molecule from the enzymes present in the cell. If the substrates for the reaction are available, the reaction will proceed and the molecule count for each metabolite participating in the reaction is updated accordingly.

### 2.3. Abstraction of folding

The protein folding process referred to in Section 1.1 provides a mapping from sequence to spatial shape. At present a computational prediction of this mapping is in most cases not practical; this is also true for the reverse mapping (Pierce and Winfree, 2002). To capture the multitude of non-local interactions among amino acid residues during the folding process, we decided to

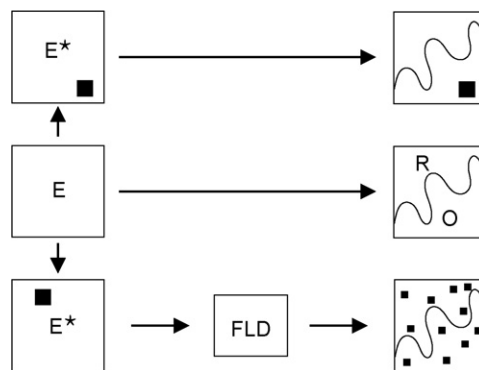


Fig. 1. Modelling of the folding process. The function of the wild type enzyme (E) can be decomposed into two conceptually independent parts, recognition (R) and operation (O); middle layer. Without the folding process a locally damaged enzyme (E\*) in the sequence is likely to affect only one of these parts; top layer. The folding process (FLD) spreads the effect of a local error (indicated as a black box) throughout the three-dimensional structure; bottom layer.

simulate protein folding in our model by means of the message digest algorithm MD5 (Rivest, 1992). The algorithm takes an arbitrarily long digital input and deterministically produces a 128 bit response that is characteristic of the input sequence. Every bit of this response can be affected by any bit of the input. The concept is illustrated in Fig. 1.

For the experiments reported here we would like to compare the three cases indicated by the three layers in Fig. 1. Such a comparison requires the output of the folding process of the wild type enzyme representation to be a valid enzyme specification, i.e. it obeys the constraint stated above. Because the MD5 process cannot be inverted, it is necessary to select enzymes that have a valid representation from a pool created by applying MD5 to random sequences. If interconnectedness is part of the selection criterion, the generation process will yield a network of enzymes specified by bit strings as described in Section 2.2. For each enzyme in the network, a bit string representation that can be used as input to MD5 in order to obtain the bit string specification of the enzyme will be known. The former is considered to be equivalent to the amino acid sequence of a protein; the latter is viewed as an analogue of the folded shape. The important property of metabolic networks thus created is that faults can be introduced into one and the same network on either sequence level or shape level. We adopt the convention that the bit string referring to the sequence is termed representation and the bit string that abstracts the protein shape is termed specification. In this sense, the left side in Fig. 1 shows representations of enzymes and the right hand side shows specifications. Only in the

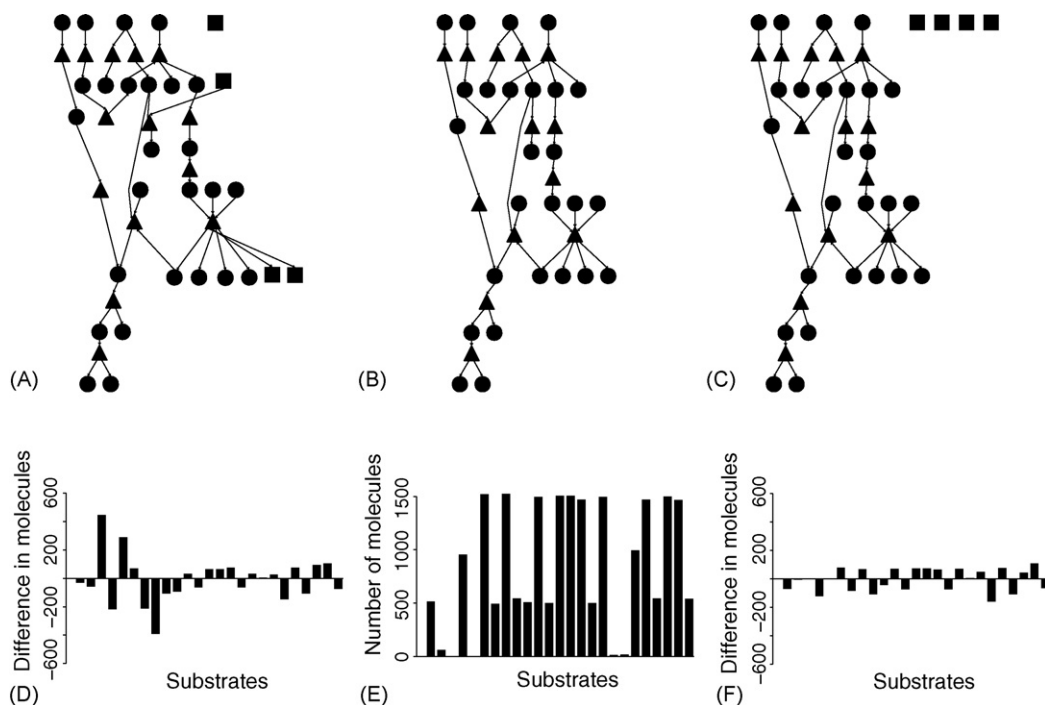


Fig. 2. Fault impact on a small sample network. Circles indicate 26 enzymes present in the wild type network, squares indicate faulty enzymes, and triangles represent substrates. The centre graph (B) represents the wild type metabolism and a snapshot of the metabolite distribution in the wild type cell (E). Without the folding process erroneous enzymes stay connected to the network (A) and the chemical composition differs significantly from that of the wild type (D). Impact of the same error probability as in (A), but with folding applied to the faulty representation, leads to the isolation of damaged enzymes (C) and concomitant small differences in metabolite concentration compared to the wild type (F).

case where folding is modelled by MD5 (the bottom row of Fig. 1), there is actually a difference between the bit string used as representation and the bit string used as specification.

### 3. Evaluation of robustness and fault impact

To quantify the impact of random modifications to an enzymatic network, we define the aberration ( $\Delta$ ) as  $\Delta = \sum_i^n (C_i^* - C_i)^2$ , where  $n$  is the number of substrates in the union of the set of substrates present in the wild type cell and the set of substrates present in the faulty cell,  $C_i^*$  the count of molecules of substrate type  $i$  in the faulty cell and correspondingly  $C_i$  is the count for the wild type cell. This measure of the impact of alterations to the cell is rather conservative as the presence of even low concentrations of chemicals which are not part of the physiological conditions of the wild type cell could be far more disruptive than what the above expression indicates. The aberration is determined at a fixed number of reaction steps after the metabolisms has been initialised. Here we use 1000 times the number of enzyme types present in the wild type cell. This insures that the initial supply of metabolites percolates

to the endpoints of the metabolic pathways. Note, that the set of substrates supplied in the reactions is always the set for the wild type cell. In other words, if a fault in an enzyme alters the substrate recognition part of the enzyme's specification to a substrate not present in the wild type cell, then this substrate will not be supplied.

A small sample network shown in Fig. 2 illustrates the topological and dynamical differences among the three cases presented in Fig. 1. After 26,000 collisions, not necessarily all resulting in reactions, the chemical composition of the well-stirred reaction medium is as shown in panel E. This serves as a reference of the wild type metabolism to which the chemical composition of networks with faulty enzymes can be compared. The fault model assumes that the bits in the enzyme type representation or specification flip with a certain probability. One or more flipped bits in the type representation or specification of an enzyme gives generally rise to a new enzyme type. For all simulations described here, the initial molecule count for any enzyme is 50. If a new enzyme type appears, 25 molecules of the new type will be added to the pool of enzymes present; no enzymes will be removed. This study focuses on the difference between

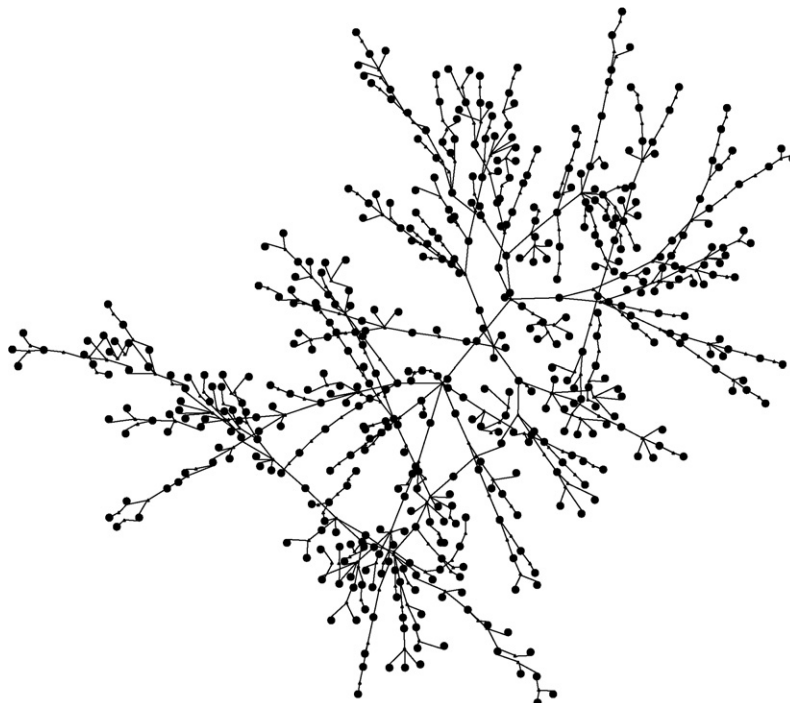


Fig. 3. Randomly generated metabolic network of 500 enzymes. Enzymes are indicated by black circles. Lines connect enzymes that catalyse consecutive reaction steps. The network contains 325 enzymes that accept one substrate, 158 of which have one product and 167 enzymes have two products. The remaining 175 enzymes accept two substrates, 87 of them generate one product and 88 generate two products. The size of substrates is minimised in this graph for clarity.

two scenarios. In the first case bits are flipped directly in the type specification bit string of enzymes. Under these conditions the enzyme may acquire a new operation or recognise a new substrate and both of these modifications are independent. The resulting network topology for an instance of this scenario is depicted in Fig. 2A. Panel D shows the difference in chemical composition that will develop with the network shown in A if contrasted with the network shown in B. Note that four new enzyme types (shown as squares) have appeared in panel A. One of them is not connected to the network; the other three are connected to the network. The second scenario assumes that faults appear in a representation that has to self-organise into the functional form. To model this situation, randomly flipped bits occur in the bit string passed through MD5 to arrive at the enzyme type specification. MD5 provides an abstraction of the coupling among all bits in the representation that would be provided by the physics of protein folding. As a consequence, local (single bit) modifications are spread in their effect globally across all of the enzyme type specification. A drastic change in the enzyme is thus very likely and therefore the faulty enzyme is not likely to be connected to the network. This is the case because the

space of possible metabolites ( $2^{18}$  with 18 bit long substrates) is much larger than the number of substrates actually present in the cell (below 1000 in the 500-enzyme network discussed below). In Fig. 2C four new types of enzymes (shown as squares) appeared, but none of them is connected. Under such circumstances, the effect on the chemical composition of the cell is minor as is evident in panel F, particularly in contrast to panel D.

We are now ready to consider a larger network. Fig. 3 shows a network of 500 enzymes that was generated with the above described process of applying MD5 to random bit strings and of selecting those that are valid and connected. This network is defined by the type bit strings of its enzymes, comprising 21,150 bit. To determine what chemical composition this network yields, 202 different substrates are provided at a level of 10,000 molecules each and the reaction dynamics is simulated for 500,000 reaction possibilities (collisions). Subsequent to the reaction dynamics typically 600–800 different metabolites are present in the cell. The chemical state of the cell after introducing random faults in the network is compared to the chemical state of the unaltered network (here called wild type), using the aberration ( $\Delta$ ) defined above. To introduce faults, it is assumed that any

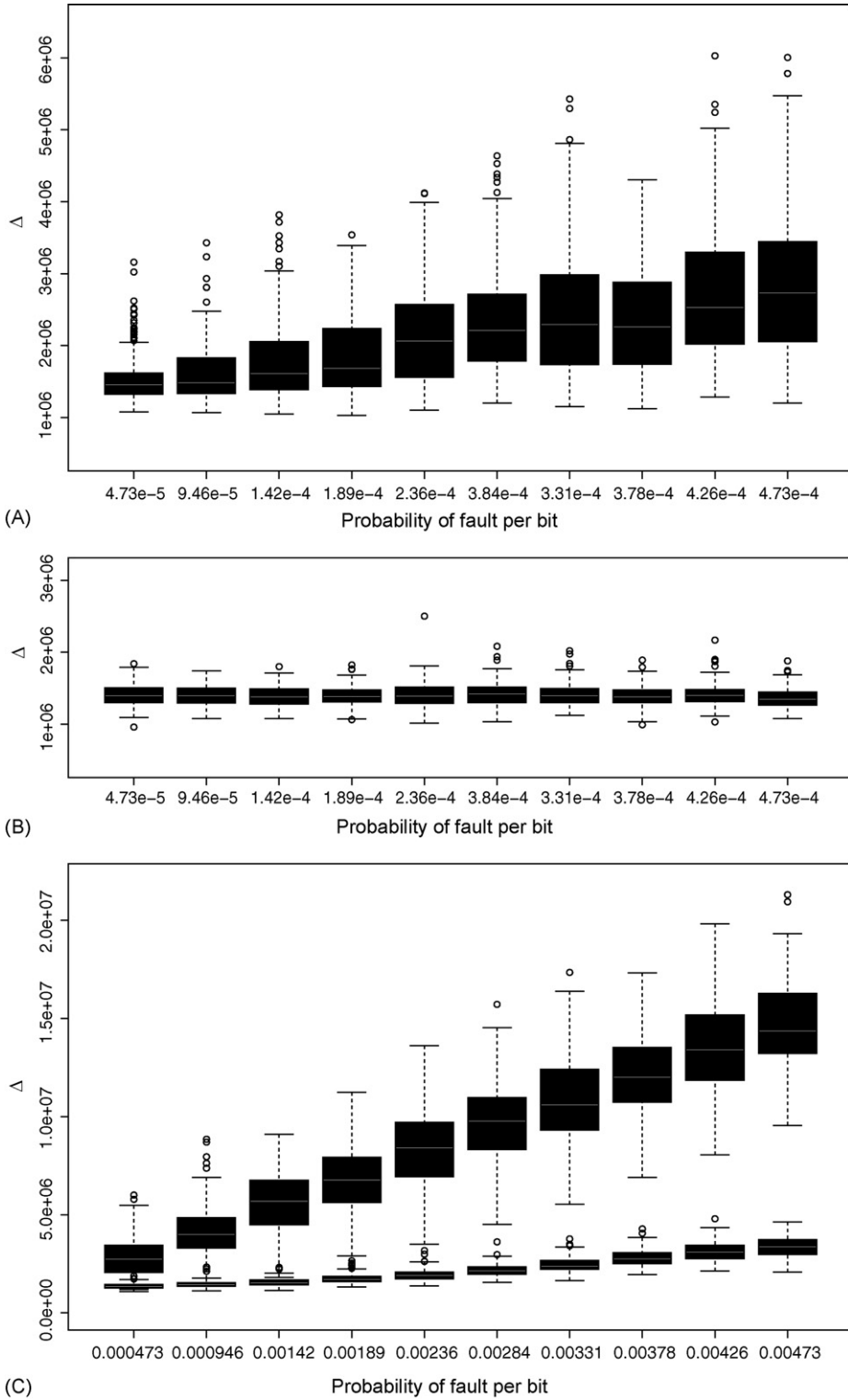


Fig. 4. Robustness of a metabolic network at low fault probabilities. Probability for a bit-flip in the enzyme representation increases from left to right. In each panel, the higher curve shows the aberration without folding, and the lower curve with folding. Mean error probabilities for the complete network are shown in (A) for 1–10 bit and (B) for 10–100 bit; 200 simulation runs were performed for each bar.

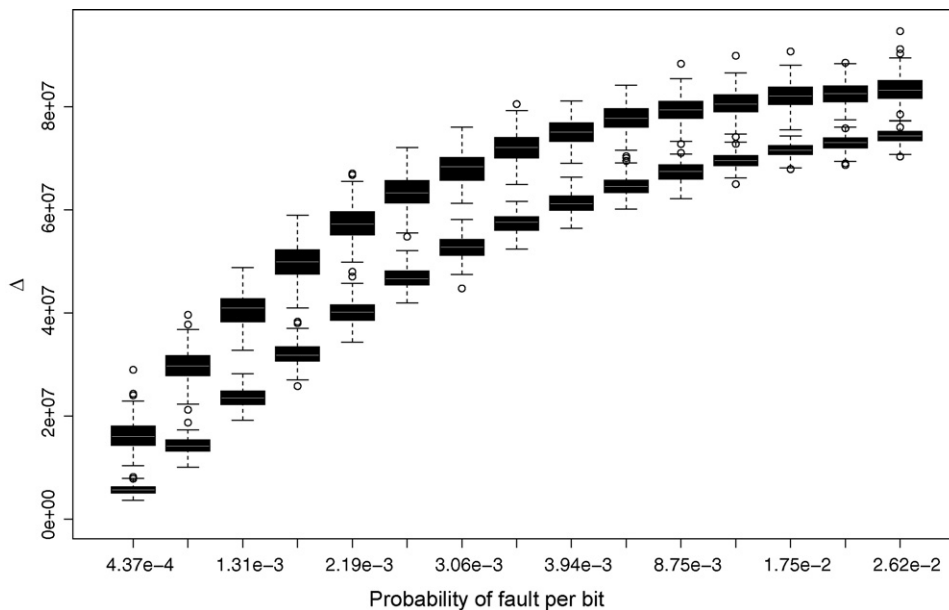


Fig. 5. Robustness of a metabolic network. Probability for a bit-flip in the 21,150 enzyme representation bits increases from a mean of 100 bits flipped to a mean of 1500 bits flipped. The bars on the higher curve aberration without folding, bars on the lower curve with folding; 200 simulation runs were performed for each bar.

of the 21,150 bits will flip with a small probability to its logic negation. We performed simulations for probabilities ranging from equivalent to a mean of a 1-bit change in the network up to a mean of 1500 bits being changed.

The impact of random faults for the two different cases we would like to contrast here is shown in Figs. 4 and 5. In the first case, the bits are flipped in the specification of the enzyme directly. In this case, no abstraction of the folding process is invoked. This case is in our model equivalent to a hypothetical direct modification of the enzymes shape by bending its surface. This could be pictured like reshaping a loose ball of aluminium foil. In the second case, the bits are flipped in the representation of the enzyme and the MD5 algorithm is used to simulate the context sensitive interactions of self-organisation that maps the amino acid sequence into a shape.

Panel A and B of Figs. 4 and 5 show how aberration ( $\Delta$ ) increases with increasing probability of faulty bits. There are two bars for each probability. For clarity in Fig. 4A the two bars are drawn on separate but identical scales. In all panels the bars located higher on the page are for the case of introducing faults directly into the specification of the enzyme and the lower bars are for the case introducing faults on the representation level. The bars extend from the first quartile (25% of sample are lower) to the third quartile (75% samples are lower) with the median marked by a white line. The whiskers show

the maximum and minimum values if they are within 1.5 times the size of the bar. Points outside this range are shown individually as circles. Each bar shows the data of 200 simulation runs, each with a different set of random faults. In this model, as the graphs show, the abstracted folding process reduces the impact of random faults on the chemical composition of the cell. The difference is most drastic at low error rates.

#### 4. Discussion

The results presented in the previous section point towards the possibility that a reduction of modularity at the component level may reduce constraints at the system level, i.e. on the phenotype. Under suitable conditions modularity in the genotype–phenotype mapping aids evolution (Jacob, 1977; Altenberg, 2004) and, conversely, interdependency among modules require more sophisticated evolutionary mechanisms (Watson and Pollack, 2005). Conrad (1977, 1998) argued that in the course of evolution gradualism, i.e. small changes in genotype lead to small changes in phenotype, is adjusted for efficient evolutionary behaviour. Our model illustrates that components highly sensitive to alterations can increase gradualism at the network level by converting interfering faults into pseudoneutral faults (Volkenstein, 1994). If the phenomenon manifest in the model is present in nature, one may find a lack of robustness at



one layer of biological organisation that serves to increase robustness at a higher level.

The primary interest here, however, is not natural evolution but the engineering of complex artificial systems. Enormous gains in raw computational power and communication bandwidth over the past decade have led to an explosion of complexity in information technology on all levels from microprocessor chips, to software systems, to globally distributed data and services. It is increasingly clear that the complexity of computing infrastructure is approaching a threshold beyond which current engineering runs into difficulties. Progress towards engineering at nanometre scale exacerbates this situation. Many engineering difficulties across these scales can be traced, at least to some extent, to a common cause: the loss of detailed prescriptive control (Zauner, 2005).

The current approach to compensate for the loss of control is the selection of functional systems from the output of the production process. To arrive at a robust complex system with an implementation paradigm based on selection is in practice at least very hard and arguably impossible. One reason is that the fraction of rejects grows fast as the component count increases. Another is that extensive testing becomes prohibitive. These two issues are interconnected by the feature of robustness. More robust architectures will provide a higher yield of functional systems, but at the same time diminish the chance of discovering problems during a non-exhaustive test. Nonetheless, these challenges are not unsurmountable as is exemplified by the marvellous mass-produced nanoarchitecture of biological cells. Self-organising processes ranging from self-assembly to growth and development will play a central role in mimicking the high integration density of cells in artificial devices. Abstracting the principles of the fabrication processes and quality control intrinsic to cells is a first step towards this goal. The results reported here hint at one such principle: redundancy combined with self-isolation of faulty components by a design that insures that faults impact interconnection. On this basis, it can be advantageous to amplify small errors to catastrophic failure. Self-organising processes, such as protein folding or development, provide the amplification in nature; the same principle can also be designed into technical systems.

The question of how one can arrive at “Reliable Organisms from Unreliable Components” (von Neumann, 1956) is as pertinent today as it was at the outset of computer science. With the benefit of half a century of insights from molecular biology, we can now contemplate the engineering paradigms necessary for narrowing the

gap in integration density between nature’s gadgets and man-made devices.

## Acknowledgements

The reported material is based on work in part supported by the University of Southampton’s Life Sciences Interfaces Initiative. The authors thank S. Dasmahapatra, R. Watson, and J. Werner for insightful discussions.

## References

- Adam, G., Delbrück, M., 1968. Reduction of dimensionality in biological diffusion processes. In: Rich, A., Davidson, N. (Eds.), *Structural Chemistry and Molecular Biology*. W.H. Freeman, New York, pp. 198–215.
- Altenberg, L., 2004. Modularity in evolution: some low-level questions. In: Rasskin-Gutman, D., Callebaut, W. (Eds.), *Modularity*. MIT Press, Cambridge, MA, USA.
- Axe, D.D., 2000. Extreme functional sensitivity to conservative amino acid changes on enzyme exteriors. *J. Mol. Biol.* 301, 585–595.
- Ball, P., 2002. Natural strategies for the molecular engineer. *Nanotechnology* 13, 15–28.
- Barabási, A.-L., Oltvai, Z.N., 2004. Network biology: understanding the cell’s functional organization. *Nat. Rev.: Genet.* 5, 101–114.
- Conrad, M., 1977. Evolutionary adaptability of biological macromolecules. *J. Mol. Evol.* 10, 87–91.
- Conrad, M., 1979. Mutation–absorption model of the enzyme. *Bull. Math. Biol.* 41, 387–405.
- Conrad, M., 1998. Towards high evolvability dynamics. In: Van de Vijver, G. et al., *Evolutionary Systems*. Kluwer Academic Publishers, Dordrecht, pp. 33–43.
- Conrad, M., Volkenstein, M.V., 1981. Replaceability of amino acids and the self-facilitation of evolution. *J. Theor. Biol.* 92, 293–299.
- Davidson, A.R., Sauer, R.T., 1994. Folded proteins occur frequently in libraries of random amino acid sequences. *Proc. Natl. Acad. Sci. U.S.A.* 91, 2146–2150.
- Dittrich, P., Ziegler, J., Banzhaf, W., 2001. Artificial chemistries—a review. *Artif. Life* 7 (3), 225–275.
- Dobson, C.M., 2003. Protein folding and misfolding. *Nature* 426, 884–890.
- Finkelstein, A.V., Galzitskaya, O.V., 2004. Physics of protein folding. *Phys. Life Rev.* 1, 23–56.
- Fischer, E., 1894. Einfluss der configuration auf die wirkung der enzyme. *Ber. Deut. Chem. Ges.* 27 (3), 2985–2993.
- Friedrich, P., 1984. *Supramolecular Enzyme Organization*. Pergamon Press, Oxford.
- Grishin, N.V., 2001. Fold change in evolution of protein structures. *J. Struct. Biol.* 134, 167–85.
- Guo, H.H., Choe, J., Loeb, L.A., 2004. Protein tolerance to random amino acid change. *Proc. Natl. Acad. Sci. U.S.A.* 101 (25), 9205–9210.
- Jacob, F., 1977. Evolution and tinkering. *Science* 196, 1161–1166.
- Kimura, M., Ohta, T., 1974. On some principles governing molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* 71 (7), 2848–2852.
- Kirschner, M., Gerhart, J., 1998. Evolvability. *Proc. Natl. Acad. Sci. U.S.A.* 95, 8420–8427.
- Lehn, J.-M., 2002. Toward self-organization and complex matter. *Science* 295, 2400–2403.

- Lindorff-Larsen, K., Rogen, P., Paci, E., Vendruscolo, M., Dobson, C.M., 2005. Protein folding and the organization of the protein topology universe. *TIBS* 30 (1), 13–19.
- Luo, D., 2003. The road from biology to materials. *Mater. Today*, 38–43.
- Luque, I., Leavitt, S.A., Freire, E., 2002. The linkage between protein folding and functional cooperativity: two sides of the same coin?. *Annu. Rev. Biophys. Biomol. Struct.* 31, 235–256.
- Netzer, W.J., Hartl, F.U., 1998. Protein folding in the cytosol: chaperonin-dependent and independent mechanisms. *Trends Biochem. Sci.* 23, 68–73.
- Pierce, N.A., Winfree, E., 2002. Protein design is NP-hard. *Protein Eng.* 15 (10), 779–782.
- Prijambada, I.D., Yomo, T., Tanaka, F., Kawama, T., Yamamoto, K., Hasegawa, A., Shima, Y., Negoro, S., Urabe, I., 1996. Solubility of artificial proteins with random sequences. *FEBS Lett.* 382, 21–25.
- Rivest, R., 1992. The MD5 message-digest algorithm. Internet Engineering Task Force (IETF), RFC, p. 1321.
- Sosnick, T.R., Berry, R.S., Colubri, A., Fernandez, A., 2002. Distinguishing foldable proteins from non-folders: when and how do they differ? *Proteins* 49, 15–23.
- Volkenstein, M.V., 1994. *Physical Approaches to Biological Evolution*. Springer-Verlag, Berlin.
- von Neumann, J., 1956. Probabilistic logics and the synthesis of reliable organisms from unreliable components. In: Shannon, C.E., McCarthy, J. (Eds.), *Automata Studies*. No. 34 in *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, pp. 43–98.
- Wang, Z., Moulton, J., 2001. SNPs, protein structure, and disease. *Hum. Mutat.* 17, 263–270.
- Watson, R.A., Pollack, J.B., 2005. Modular interdependency in complex dynamical systems. *Artif. Life*. 11 (4), 403–405.
- Zauner, K.-P., 2005. From prescriptive programming of solid-state devices to orchestrated self-organisation of informed matter. In: Banâtre, J.-P., et al. (Eds.), *Proceedings of the UPP 2004*, vol. 3566 of LNCS. Springer, Berlin, pp. 47–55.