Chapter 11
Design of Experiments

Abstract. Latin squares and hypercubes are obtained as solutions of the wave equation.

Multivariate modeling potential for evolvable designs of experiments is evaluated.

The general PSM framework is presented as flexible guideline for a large variety of designs of experiments.

Case studies refer to pharmaceutical pipeline, to drug discovery and development and to printed circuits quality evaluations. New informational entropy criteria have been applied for 2-phenylindole derivatives library design.

11.1 Convection Model

The connection with designs of experiments was established and illustrated in the study of matrices generations flowsheeting and coupling of operations (Iordache 2009, 2010).

Consider the convective part of the first-order wave equation, WE (eq. 3.4, eq. 3.5):

$$\frac{\partial Y}{\partial T} \oplus V \otimes \frac{\partial Y}{\partial Z} = 0 \tag{11.1}$$

The initial condition is:

$$Y(Z, 0) = F(Z) \tag{11.2}$$

The operations are the sum $\oplus$ and the product $\otimes$ in GF (m).

The general solution of the partial first-order wave equation, WE, is:

$$Y(Z, T) = F(Z \oplus (V \otimes T)) \tag{11.3}$$

Consider the initial condition:

$$Y(Z, 0) = F(Z) = Z \tag{11.4}$$

This means that at T=0, the output Y of the classification schema at the distance Z in schema is exactly Z. This schema is that in which each new classification level activates a new difference in properties allowing classification. The initial condition ensures that the wave of the classification or separation process is initiated and is going on.

It results in the characteristic:

$$Y = Z \oplus (V \otimes T) \tag{11.5}$$

The GF (3) solution is presented in detail. For T=0 the solution Y is shown in Table 11.1.

Table 11.1 Convection model, m=3: Y (0, Z)

Z\V	0	1	2
0	0	0	0
1	1	1	1
2	2	2	2

For T=1 the solution is shown in Table 11.2.

Table 11.2 Convection model, m=3: Y (1, Z)

Z\V	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

For T=2 the solution of the convection model is shown in Table 11.3:

Table 11.3 Convection model, m=3: Y (2, Z)

Z\V	0	1	2
0	0	2	1
1	1	0	2
2	2	1	0

There is a relation between different solutions of the first-order wave equation, WE, and conventional DOE matrices.

For different values of T, T=1, T=2 one obtained different (3x3) Latin-squares.

Latin squares close association to DOE is well known (Hedayat et al. 1999)

The procedures to obtain DOE are suggested by universal constructions in categorical framework. There are several DOE to be obtained by combining the solutions obtained for different values of T.

Table 11.4 Concatenated solutions, m=3

000	012	021
111	120	102
222	201	210

Superposing by concatenation the elements of the Table 11.1, Table 11.2, and Table 11.3, Table 11.4 will result.

Table 11.4 shows the concatenated solutions, for m=3.

Pasting down the 3-digit numbers from Table 11.4, column after column, Table 11.5 is obtained.

Table 11.5 contains the pasting-down columns.

It is a DOE with nine experiments for three factors, denoted here F0, F1, and F3.

Table 11.5 Pasting down columns, m=3

F0	F1	F2
0	0	0
1	1	1
2	2	2
0	1	2
1	2	0
2	0	1
0	2	1
1	0	2
2	1	0

Columns in Table 11.5 are orthogonal. Each column corresponds to first-order wave equation, WE, solutions at different velocities V. Associating one supplementary digit for each column in Table 11.4, the four-digit numbers as in Table 11.6 result. Here (0) is associated to the first column in Table 11.4,(1) to the second column and (2) to the third column.

The resulting 4-digit numbers from Table 11.6 correspond to columns of well-known orthogonal design with 9 experiments and 4 factors (Taguchi 1986, 1987, Hedayat et al. 1999).

Table 11.6 shows the indexed concatenated solutions for m=3.

Table 11.6 Indexed concatenated solutions, m=3

(0)000	(1)012	(2)021
(0)111	(1)120	(2)102
(0)222	(1)201	(2)210

Concatenation and pasting-down operations are related to the coproduct "$\cup$" type of operation in categorical framework.

The previously obtained matrices are linked to the tensor product interpretation as coproduct "$\cup$". Obviously making use of tensor products as categorical product "$\times$" will give another class of solutions, asking for significantly more experiments.

11.2 Latin Hypercubes

Computer experiments are widely used for the design and development of products.

An actual reason for promoting the use of computer experiments is that physical experimentation is maybe expensive or out-of-the-way. Latin hypercube designs are beneficial for space-filling capability (Cioppa and Lucas 2007).

A class of orthogonal Latin hypercubes obtained by Ye will be presented as WE solution (Ye 1998).

The products we use, Kronecker or Hadamard, are necessary to reduce the dimensionality for data (Kolda and Bader 2009).

The construction is based on three matrices denoted by S, M and T (Ye 1998, Nguyen 2008).

The matrix S shown in Table 11.7 is obtained as solution of the kinetic part of the WE (section 3.3).

Table 11.7 Kinetic model: Matrice S

1	1	1	1
1	-1	1	-1
1	1	-1	-1
1	-1	-1	1

The matrix S is a Walsh-Hadamard design.

The matrix M shown in Table 11.8 is obtained as solution of the convection part of the wave equation WE.

Table 11.8 Convection model: Matrice M

1	2	3	4
2	1	4	3
3	4	1	2
4	3	2	1

The matrix M is a Latin square.

The matrix T as shown in Table 11.9 results using Hadamard product of S and M.

Table 11.9 Hadamard product: Matrice T

1	2	3	4
2	-1	4	-3
3	4	-1	-2
4	-3	-2	1

The Latin hypercube results by the method described in the literature (Ye 1998), Cioppa and Lucas 2007).

Table 11.10 shows the Latin hypercube.

This consists in pasting down two times the matrix T with an intercalated null row.

Table 11.10 Latin hypercube

1	2	3	4
2	-1	4	-3
3	4	-1	-2
4	-3	-2	1
0	0	0	0
-1	-2	-3	-4
-2	1	-4	3
-3	-4	1	2
-4	3	2	-1

The wave equation appears to be the source of new classes of orthogonal Latin hypercubes.

11.3 Self-Evolvable DOE Frames

The design of experiments for the exploration of high-dimensional experimental spaces may be addressed by evolvable DOE and EDOE methods.

Significant applications concern drug discovery.

We will consider that the activity of new drug discovery can be divided into four basic modules or steps.

The first module K0 corresponds to resources and to research step.

The second module K1 should be based on designs of experiments, DOE.

The third module K2 is a meta-design and for this reason it was denoted by 2-DOE.

The fourth module K3 is a meta-meta-design and for this reason may be denoted by

3-DOE.The general method is illustrated in Fig. 11.1.

The four modules of variation, K0, K1, K2 and K3, are denoted also by S, 1-DOE, 2-DOE, and 3-DOE.

To start the EDOE, we examine experimental space of properties.

After a number of iterations at this level, we may make predictions of drug-likeness too.

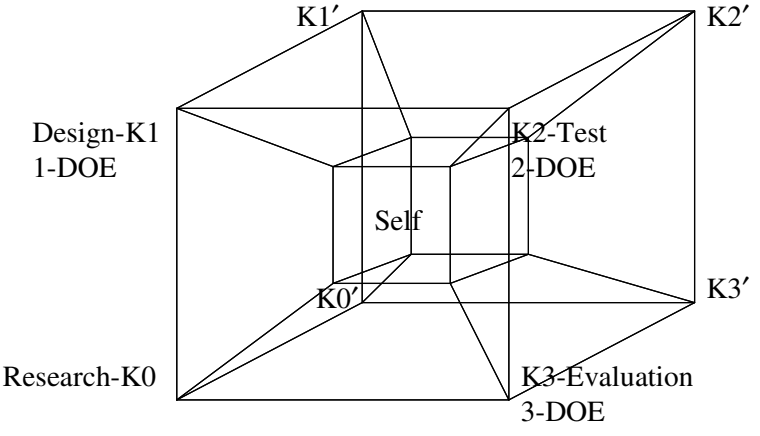

Fig. 11.1 Polytope for self-evolvable DOE basic framework

The notations are: K0-Research, K1-Design, 1-DOE, K2-Tests, 2-DOE, and K3-Evaluation, 3-DOE.

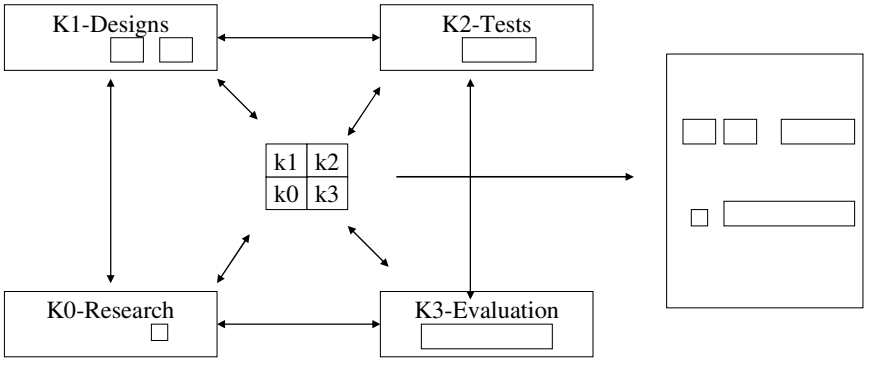

Fig. 11.2 Operad for self-evolvable DOE framework

Fig. 11.1 suggests that after the integration way we need to look at the differentiation way. This may consist in changing the hierarchical orders in the matrices associated to different DOE. Making use of the developments of the direct way may result in a kind of symmetry-breaking result for the reverse way. The swinging from direct to reverse investigation is beneficial for new designs testing because the boundaries where creative research grows and new information is created consist of synchronized integration and differentiation tendencies.

Fig. 11.2 highlights the operadic aspects of the self-evolvable DOE framework.

The DOE may be associated to a set of molecules or an embedded design if a genomic analysis is possible. The fourfold framework may be applied to just one of the levels.

Table 11.11 summarizes the categorification steps for DOE

Table 11.11 Categorification for DOE

Level	K0	K1	K2	K3	Self
-	n=0	n=1	n=2	n=3	n≥4
Categories	0-category	1-category	2-category	3-category	4-category
Example	Research	DOE	2-DOE Tests	3-DOE Evaluation	Self Evolvable

The study of DOE for self-integrative closure and the emergence of self-evolvable DOE systems corresponding to n≥4 represent a challenge.

Let us restrict here to the discovery stage associated to K1 as a first example.

Different classes of Latin square designs may be obtained as solutions of the wave equation.

A method of designing chemical substances was presented by Wood and Rose (1999).

The method allows sampling combinatorial chemistry space for synthesis based on DOE with Latin squares or more general with orthogonal arrays.

Libraries with four sites of variation for molecules may be designed using Greco-Latin squares.

Consider four sites of variation, k10, k11, k12 and k13 for substitute groups. They correspond to sub-levels of the level K1.

Then only four different substitutes are selected for each substitute or pendant group, k10, k11, k12 and k13.

The substitute group k10 consists of four candidates, denoted 1, 2, 3 and 4, the substitute k11 from four candidates denoted a, b ,c ,d, the substitute k12 of four candidates denoted A, B, C, D and substitute k13 of four candidates denoted α, β, γ, and δ.

Recall that the wave equation is able to generate Latin squares as solutions if the algebraic structures of functions and parameter are Galois Fields (Iordache 2009, 2010).

Superposition of such solutions of the wave equation gives Greco-Latin squares as shown in Table 11.12. This superposition represents a specific categorical product.

Table 11.12 shows the matrix of a Greco-Latin design.

For this table the sub-levels of the level K1 are: k10 = {1, 2, 3, 4}, k11 = {a, b, c, d}, k12 = {A, B, C, D}, and k13 = {α, β, γ, δ}.

Table 11.12 Greco-Latin square design

k10\k11	a	b	c	D
1	Aα	Bβ	Cγ	Dδ
2	Bδ	Aγ	Dβ	Cα
3	Cβ	Dα	Aδ	Bγ
4	Dγ	Cδ	Bα	Aβ

With 16 experiments only we may obtain significant information.

Running the experiment, we may select the critical substituents.

EDOE framework should be seen as a general pattern rather than as a predetermined fixed plan. This means that we may have partial Greco-Latins as micro-arrays.

For different values of time T in the solution of the convection part of the wave equation, WE, we may obtain all the orthogonal Latin squares.

A superposition of this mutually orthogonal Latin square gives the so-called Trojan squares. The presentation as Latin cube is shown in Fig. 11.3.

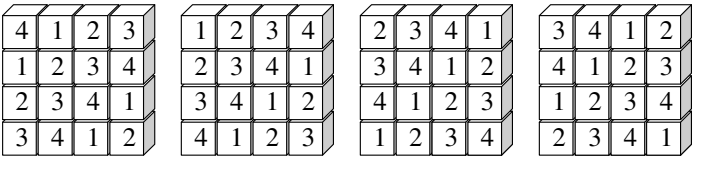

Fig. 11.3 Latin cube of order four

The wave equation may generate inflated Latin squares as solutions if the algebraic structures of functions and parameters are cyclic groups. Such modified Latin squares have been studied by Bailey (Bailey 1992). Consider that the sub-levels of K1 are the same.

The cyclic-group-based solutions of the wave equation give inflated Latin squares shown in Table 11.13. The inflation which replaces each letter by four new letters corresponds to another definition of the categorical product.

Table 11.13 Inflated Latin square design

1aAα	2bBβ	3cCγ	4dDδ
4dDδ	1aAα	2bBβ	3cCγ
3cCγ	4dDδ	1aAα	2bBβ
2bBβ	3cCγ	4dDδ	1aAα

Table 11.14 Comparison of DOE and self-evolvable DOE

Characteristics	Conventional DOE Robust design	Self-evolvable DOE
Resources	Considered very large	Definite resources-time frames
Objectives	One solution, one goal, mono-centric, search for ideal optimal solutions	General goals, multi-centric, workable, acceptable, viable solutions
Strategies	Prescribed, fixed	Self-adaptive, self-evolvable
Matrices	Fixed matrices–specific, developed Defined-based on learning Conventional design-detailed models	Constructed, self-adaptive Sensitive to case study, undefined Generated by WE and evolved
Methods	Single DOE, Sequential	Frameworks of small DOE Parallel and sequential
Required data	Clear steps, complete data	Incomplete data and variables
Factor hierarchy	Ignored; Pareto charts; inner-outer factors	Centered-heterarchy, self-evolvable frames
Approach	Mainly deductive Isolate from medium	Deductive and inductive Builds, dissociates, reconstructs
Interactions	Typically neglected, linear	Considered from start, useful
New experiments	Controlled use, confirmation experiment	Self-adaptive, evolution, minimal
Selectivity	Selective, robust to noisy	Robust to redundant information and noisy
Criteria	Average, variance, signal/noise	Informational entropy, distance
Calculus	Real field	Real and other than real fields
Data analysis	Developed-completed	Self-adaptive, interactive, online
Learning	Early-learning	Beyond learning, creative, self-evolving
Time concept	Neglects time or considers linear time only	Time scales, cyclic time frames
Complexity	Considered as enemy, avoided	Considered as ally, accepted and utilized

The complete EDOE frame implies to continue the cycle from DOE matrix that is from K1 level, to tests that is K2 level, evaluation and implementation that is K3 level as shown in Fig. 11.1.

Table 11.14 compares conventional DOE to self-evolvable DOE methods.

11.4 Multi-level Data Analysis

A structural model pertaining hierarchical classes analysis is presented in the following. The case study refers to solderability and surface finishes. Solderability is the ability of a surface to be wetted by molten solder.

Good solderability for the PCB, as well as for the components has become an important element in achieving the quality required in competitive markets. Microelectronics requires the production of reliable assemblies in an ultra-low volume environment. As the new assembly technologies, such as ball grid array, flip chip, and chip on board, have progressed, there has been a demand to obtain new solderable surface finishes, alternative to the conventional hot-air level soldering, HASL. Complex PCBs demand to increase the functionality of the final surface finish.

The challenge for printed circuits industry is to correlate the surface finish technology to specific application.

Table 11.15 shows the surface finish quality framework.

Table 11.15 includes the main factors for the surface finish quality (Iordache 2009).

They pertain to design, materials, processes and applications.

Table 11.15 Surface finish quality framework

Design, D	Materials, M	Process, P	Application, A
D1-Pad size	M1-SF type	P1-SF thickness	A1-Therm. Cycling
D2-Hole size	M2-SM type	P2-SM application	A2-IST
D3- Heat transfer	M3-Flux type	P3-Contamination	-
-	M4-Solder type	-	-

SF-surface finish, SM-solder mask, IST-interconnect stress test.

The surface finish selection system is based on EDOE methodology. The center DOE contains PCB design factors, materials factors referring to surface finish and its application for solderability, processing factors and testing factors referring to solderability after reliability tests (thermal cycling and IST). DOE steps alternate with measurements and analyze steps followed by DOE reorganization.

Different tests may be performed to evaluate surface finish, SF, solderability.

Examples of typical tests are:

- MUST- wetting balance test
- SERA- sequential electrochemical reduction analysis
- Dip & Look-standard solderability test.

The problem is to quantify the results of all these solderability tests. An example of global criterion that summarizes the partial test significance is introduced next. It associates a real value to a digitalized vector and allows subsequent modification of the DOE matrix.

Denote by S= [i_1,...,i_k...] the test resulting vector for a sample. Here i_k is the digit "1" or "0" corresponding respectively to the result pass or fail of the kth test. To any global test result S, a valuation V(S) defined by: $V(S) = \Sigma_k\ i_k\ (0.5)^k$ is associated. According to the valuation formula, the solderability for any sample depends on the significance associated in the hierarchical testing sequence to partial tests. For this case study, the first test has a weight 0.5, the next 0.25, the following 0.125 and so on, the proposed hierarchy for tests being: MUST>SERA>Dip& Look>….

The valuation V(S) is in fact a similarity as defined in Section 2.3. V(S) gives similarities relative to a reference vector containing only "1" as coordinates.

The experiment is based on EDOE methodology. DOE steps alternate with measurements and analyze steps followed by DOE reorganization.

Table 11.16 shows the element of surface finish-Design-D

Table 11.16 Surface finish-Design-D

Factor	"-1"	"1"
D1-Pad size	Small	Large
D2-Hole size	Small	Large
D3-Heat transfer	With	Without

The test mini-coupon allows performing all the solderability tests after an imposed number of interconnect stress test, IST, cycles. It evaluates solderability as a function of testing time.

Fig. 11.4 shows the modules of self-evolvable DOE frame for surface finish.

Fig. 11.4 suggests that we need to look at the integration way, K0→K1→K2→K3 and at the differentiation way, K3′→K2′→K1′→K0′.

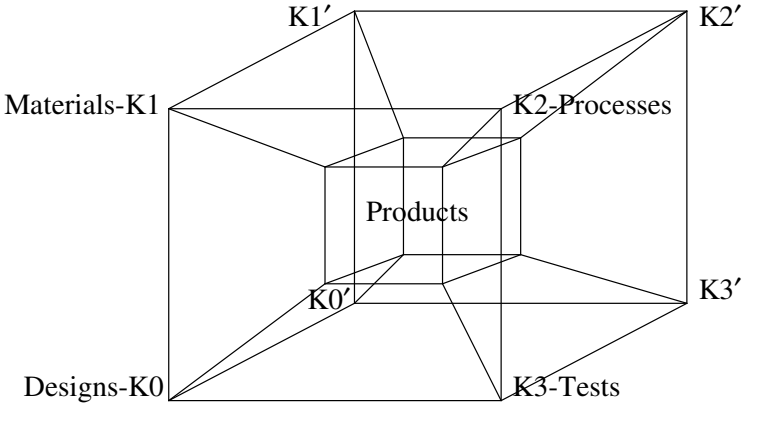

Fig. 11.4 General self-evolvable DOE frame

Preliminary tests assured that materials-M and processes-P factors are significant variables. Table 11.17 contains the notations for materials-M settings. ENIG denotes electroless nickel, immersion gold finish. S/M type depends on supplier (“D”, “C” or “T”).

Table 11.17 Surface finish-Materials-M

Factor	“-1”	“0”	“1”
M1-Surface finish	ENIG(P)	Electroplate Ni/Au	Ag
M2-Solder masks	“D”	“C”	“T”

The DOE matrix of type L 9,2,3 with three settings is considered (Iordache 2009). The main factor is M2 that is the S/M type. The processing-P factors are included in Table 11.18.

Table 11.18 Surface finish-Processing-P

Factor	“-1”	“1”
P1-Coating thickness	High	Low
P2-SM application	After	Prior

From the DOE it results that the factor P1, the coating thickness, is more significant for solderability. An interaction experiment for M, P and A was performed at this stage on the first level of EDOE. The electroless Ni/Au and a compatible mask indexed by “T” have been selected to perform an application test. Nine values of the thermal cycling parameters corresponding to the application test have been considered. The time step for the number of cycles is 25 cycles. In this case -4 is linked to 0 cycles, -3 to 25 cycles, and so on till +4 that is linked to 200 cycles. The corresponding matrices are of the type Ln,m,s (Iordache 2009).

Table 11.19 shows the factors for materials-M, processing-P, and application-A.

Table 11.19 Factors for materials-M, processing-P, application-A

Factor	-4	-3	-2	-1	0	1	2	3	4
A1-IST cycle	0	25	50	75	100	125	150	175	200
Factor	“-1”			“0”			“1”		
M2-SM type	“D”			“C”			“T”		
P1-Coating thick.	High			Avg.			Low		

The M, P, A factors are lumped together in the interaction type of experiment from Table 11.19. This test shows that the application factor that is thermal cycling plays significantly. The set of resulting DOE matrices is useful in the implementation of new surface finish technology.

The EDOE method allows simulations for new surface finish.

New vectors, that is, settings of parameters for an experiment may enter as new rows in the DOE matrices. New vectors have been denoted as italicized rows.

Suppose for instance that the new processing vector will be: [*1, -1*]. This designation is followed by a forward step. The performed experiment will be classified as run 2 in processing-P matrix. This allows predicting solderability valuation of 0.75. Consider also a new designation step in which the design-D vector is [*1, 1, -1*]. This is classified in the same class as the first run in matrix D. If high solderability means valuation higher than 0.75 these two vectors that is new experiments will provide digits "1" in the center matrix. These replacements summarize the information and represent backward steps. They translate a real valuation of the solderability into a digit only. Observe that a calculus of valuation that is of similarity allows shifting from forward to backward steps.

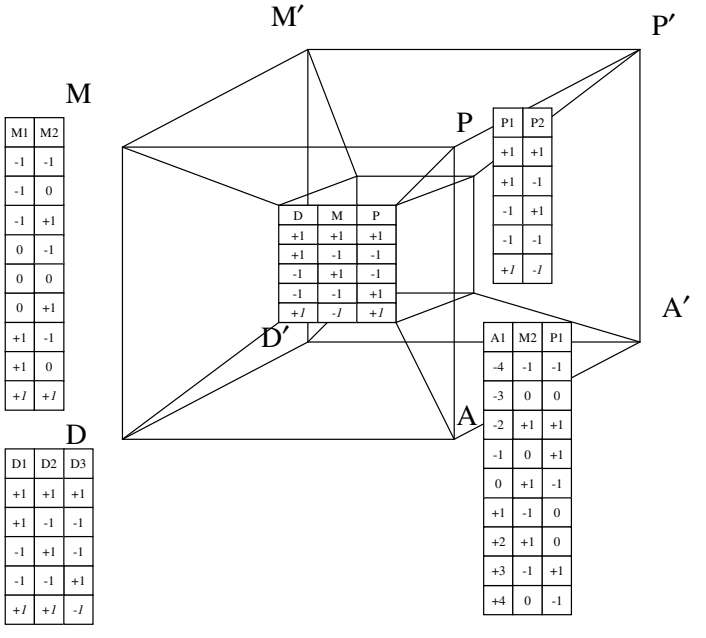

Fig. 11.5 Self-evolvable DOE frame for surface finish

Coupled with materials M corresponding to lower than 0.75 solderability (that is "-1" digit in the central matrix) the new vector in the central matrix will be [*1, -1, +1*]. This is similar to the second run [1, -1, -1] in the initial center design and it is predicted that it will show performances similar to that run.

Fig. 11.5 suggests that after the integration way D→M→P→A we need to look at the differentiation way A′→P′→M′→D′. This may consist in changing the hierarchical order or the elements in the matrices associated to D, M, P and A, in removing less significant factors. Making use of the developments of the direct way will result in a kind of symmetry-breaking result for the reverse way. The swinging from direct to reverse investigation is beneficial for self-evolution.

11.5 Pharmaceutical Systems

Designing, building and controlling complex systems became a central challenge for scientists and engineers in the coming years. A new approach to problem solving for complexity is represented by the evolvable designs of experiments, EDOE (Iordache 2009). It is based on the thesis that knowledge cannot be a passive reflection of reality, or a passive application of a formal problem-solving model, but has to be more of an active and interactive construction. EDOE is a modern way to cross industrial and technological complexity frontiers by replacing pre-programmed and fixed designs and problem-solving methods by evolvable ones.

The EDOE methodology may find applications for complex problems as the so-called pharmaceutical pipeline.

This refers to the new product, to research and development in pharmaceutical industry.

The typical sequence for new product implementation contains the following main steps:

Resources →Discovery→ Development→ Launching

Biological, chemical and other resources allow the discovery of drug lead.

The development step includes tests, preclinical, P0, followed by three phases of tests, PI, PII, and PIII.

The product launching starts with NDA, New Drug Application, and FDA, Food and Drug Administration, submissions and reviews, and continues with production and marketing steps.

Some areas of pharmaceutical industry are facing a productivity crisis (Woodcock and Woosly 2008). Despite rising investment in pharmaceutical research and development, successful development of new drugs is slowing. The high costs of new drugs development may discourage investment in more innovative, risky approaches in therapeutics.

The FDA, with its dual role of promoting and protecting health is charged with implementing policies that ensure that the benefits of the new products will surpass their risks, while simultaneous by promoting innovations that can improve health.

It was observed that chemical and biological systems may have huge behavior spaces and laboratory experiments and models cover only tiny aspects of a system's behavior.

The models often ignore the essential temporal and conceptual space organization of the research and implementation components. Moreover, models and methodologies lack flexibility to adapt and to faster represent more areas of the behavior space.

They neglect synergies – beneficial, nonlinear interactions between systems that cannot be inferred from existing resources and may be missed.

The architecture of the models should be in correspondence with that of the studied system within physically, biologically or cognitive recognizable spaces.

This will require combining multiple-level modeling methods in innovative ways, multiple levels of organization activated both in parallel as in series.

It is a need for new modeling and simulation methods, sufficiently flexible, adaptable and evolvable that is able to explore larger portions of the behavior space, a strong request for cognitive architecture reflecting the essential temporal and spatial organization of the real substrates and allowing autonomy of the new product development system.

PSM and more specifically EDOE, are promising cognitive architectures proposed as new methodologies for self-level problem solving in pharmacology.

The PSM general framework is based on four modules and their self-integrative closure.

Fig. 11.6 suggests a transition from the pharmaceutical pipelines to pipecycles.

The module K0 corresponds to substrate and resources, the module K1 to discovery step, K2 to developments and tests and the module K3 to product implementation and launching.

The first module involves resource mining. Resources are material, biological and of knowledge type.

The second module K1 is that of discovery and involves in this case drug-like molecules discovery, lead discovery and optimization. It may be a DOE.

The third module K2 is that of drug testing and development. It is a meta-design and for this reason may be denoted by 2-DOE since refers to processing DOE.

The fourth module K3 includes application and approval processes, manufacturing, marketing and monitoring of the product.

Each module may involve several sub-modules organized as epicycles.

For instance, in the module K2 there exists a natural cycle P0, P1, P2 and P3.

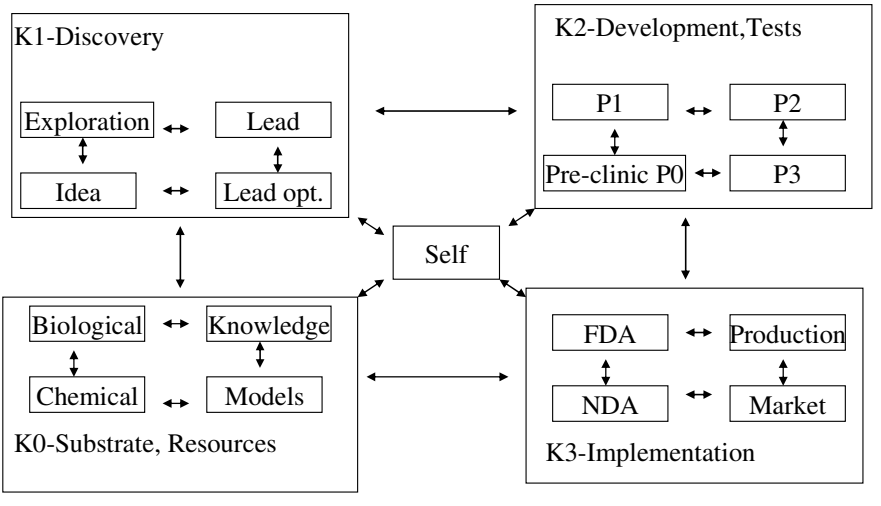

Fig. 11.6 Pharmaceutical pipecycles

For the module K3 the NDA step is followed by FDA step this by production and this by product marketing.

The transition from pipeline to pipecycles proposes a methodology that closes the loop in iterated experimentation in a high dimensional space. The cycling refers to large cycles for the whole process of four modules or just to one module or sub-module and the corresponding epicycles.

Some cycles may be fully automated if autonomous experimentation methods are used to conduct high-throughput experiments.

Modeling of matrix designs and use of informational criteria accelerate the development of new drugs.

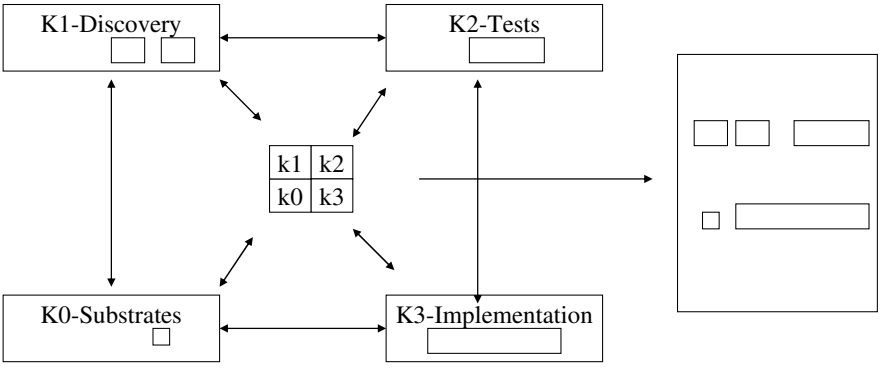

Fig. 11.7 Operad for pharmaceutical pipecycles

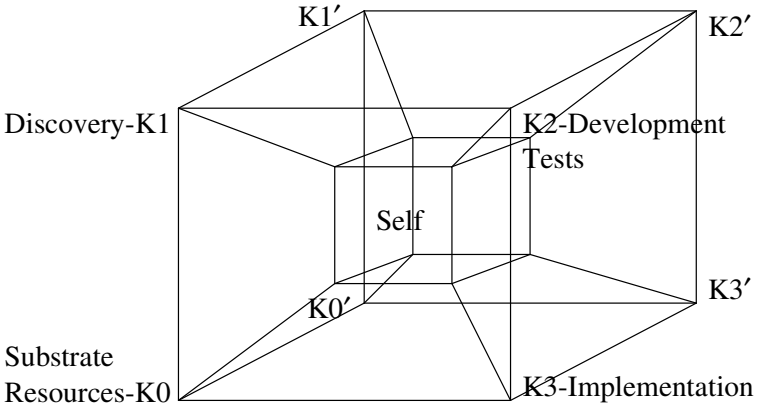

Fig. 11.8 Polytope for pharmaceutical pipecycles

Fig. 11.7 shows the operad for pharmaceutical pipecycles.

K0-Substrate, Resources, K1-Discovery, K2-Development Tests, K3-Implementation.

Fig. 11.8 illustrates the polytope of pharmaceutical pipecycles.

There exist an integration way and a differentiation way. The reconciliation of these two ways is a negotiated process since lateness in drug discovery should be avoided.

11.6 Library Design by Entropy Criteria

Applicability of PSM methodology for 2-phenylindole library design is described here.

Some 2-phenylindoles and their derivative prove to have anti-cancer activity (Basak et al. 2010).

Fig. 11.9 shows the molecular structure of the 2-phenylindole derivatives. Different radicals are denoted by R1, R2, R3 and X (see Table 1, Basak et al. 2010).

For 2-phenylindole derivatives we considered associated vectors as: y= [R1, R2, R3, X]

This means that the chosen significance order is R1>R2>R3>X.

We associate the digit "1" to R1=H, R2=H, R3= OCH_3 and X= $C(CN)_2$ and the digit "0" to the radicals that are different from these.

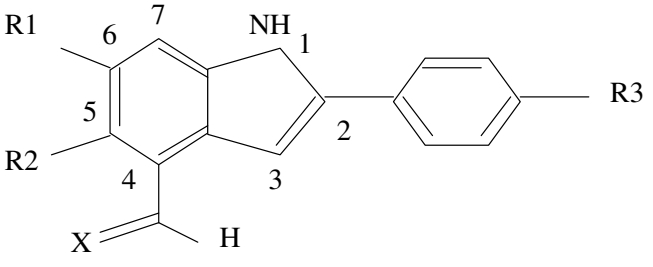

Fig. 11.9 Molecular structure of 2-phenylindole derivatives

We select the compound #2 (Table 1, Basak et al. 2010) as reference #2 = [1 1 1 1 1]

Obviously the choice of the set of characteristics and of their hierarchy should be based on drug physiological mechanism and drug efficiency tests.

Table 11.20 shows the radicals pertaining to different 2-phenylindoles derivatives.

Table 11.21 outlines the reference set for 2-phenylindole derivatives-matrix

Table 11.21 contains the same information as Table 11.20 in digitalized form.

The half maximal inhibitory concentration (IC_{50}) is a measure of the effectiveness of a compound in inhibiting biological or biochemical function. This quantitative measure indicates how much of a particular drug is needed to inhibit a given biological process by half. Here we focused on derivatives with high IC_{50}.

Table 11.20 Reference set for 2-phenylindole derivatives

No	R1	R2	R3	X	IC_{50}(nM)
#2	H	H	OCH_3	$C(CN)_2$	720
#43	H	H	H	O	420
#30	H	t-Bu	OCH_3	O	280
#19	H	n-Hexyl	CF_3	$C(CN)_2$	150

Table 11.21 Reference set for 2-phenylindole derivatives-matrix

No	R1	R2	R3	X
#2	1	1	1	1
#43	1	1	0	0
#30	1	0	1	0
#19	1	0	0	1

The compounds of the reference set have been selected to obtain a Walsh-Hadamard matrix for DOE. This offers the necessary variability for all types of drug compositions and substrate interaction.

Adding a new compound to this reference mixture, the entropy H varies.

There is only a small change of entropy, ΔH if the vector of the test compound is similar to the reference set and this supplementary compound is thought to have similar properties.

If a database 2-phenylindoles shares similar bit patterns with reference set molecules, adding a similar compound will induce a change targeting the minimum entropy production.

By contrast, inclusion of 2-phenylindoles derivatives compound having dissimilar vector leads to a higher entropy production, targeting the maximum entropy production, MEP criterion.

In this way database compounds may be screened to identify compound that causes low or high changes of the reference set informational entropy and detects other promising drugs according to the established goal.

Table 11.22 Informational entropies for 2-phenylindoles

Mixture	Reference	Reference #11	Reference #24	Reference #33	Reference #28
Matrix	[1111] [1100] [1010] [1001]	[1111] [1100] [1010] [1001] [1011]	[1111] [1100] [1010] [1001] [0110]	[1111] [1100] [1010] [1001] [0011]	[1111] [1100] [1010] [1001] [1000]
H	7.5418	11.2615	12.8343	12.8343	11.2615
ΔH	0	3.7197	5.2925	5.9525	3.7197
DD	0	0.6348	0	0	0.6348

The component #11= [1011], #24 = [0110], #33= [0011], #38= [1000] were tested since they show high (IC_{50}).

The results are shown in Table 11.22.

Table 11.22 shows the informational entropies for 2-phenylindoles.

H denotes the entropy associated to the matrix shown in first row.

ΔH denotes the difference between the entropy associated to reference and the entropy associated to reference plus one new compound.

DD denotes the distance between the reference and the matrices corresponding to reference, plus one new compound, as shown in successive columns of Table 11.22.

It appears that supplementing the reference mixture by #11, or #38, has lower effect for entropy than # 24 or #33.

The compound #11 or #38 may be preferred for a conservative new drug search based on similarity and #24 or # 33 for an innovative search based on dissimilarity.

High DD corresponds to differences allowing multiple, different classes and potential versatility of interaction. It was associated to maximum production of entropy production, MPEP criterion.

To illustrate the selection criterion at this level we take into account that the organisms varies and show biorhythms. For different regimes for organism the delivery of different 2-phenylindole mixtures may be beneficial and ensures the evolvability maximization, EM that may be evaluated by comparing DD values.

The DD criteria suggest using reference, #11 or #28, for maximum activity periods and #24 or # 33 for minimum activity periods.

EM and SEM criteria should be correlated with the methods to monitor the biomarkers of the periodic functioning of organism (Ashdown 2004, Coventry et al. 2009). Researchers have discovered that the body's immune system can destroy some cells within a window occurring every 12 to 14 days. By giving low-dose treatment at exactly the right time, they succeeded in halting the spread of advanced disease. Also they found the body has an immune cycle during which it swings "on" and "off". When the immune system turns off, it releases inhibitory cells which prevent it from fighting the disease. Treating organisms at the right time may maximize their evolvability. The timed drug delivery supposes an iterated screening of drugs and drug delivery by interaction with the organism, resources and environment.

Chronotherapy, which is an optimization of dose-time medication schedule, has been successfully applied for decades. The effects of chemotherapy exhibit circadian rhythms since the proliferation of normal cells and of damaged cells is gated by the circadian clock, damaged cells being less well synchronized. It is also known that the detoxification of cytostatic drugs depends on time of administration.

11.7 Self-Evolvable Experimentation Systems

Quantitative, predictive understanding of complex systems requires adequate information. High-throughput methods and laboratory automation technology have the potential to deliver the necessary data. To harvest this potential, experimental systems have to become evolvable and autonomous.

Self-evolvable experimentation systems are computational systems capable of autonomously investigating large experimental parameter space (Matsumaru et. al. 2004, Lovel and Zauner 2009).

Such systems should develop hypotheses, plan experiments and perform experiments in a closed-loop manner without human interaction.

Fig. 11.10 illustrates the autonomous experimentation architecture principle.

The notations are: K0-Experiment, K1-Model, K2-Prediction, and K3-Fitness.

It is a self-integrative closure technique.

The levels may be identified as follows: K0-Experiment, K1-Model, K2-Prediction, K3-Evaluation and Fitness. The model is empiric. To these the central level of self-evolution linked to the previous levels is joined. The center is considered either as the starting area or as the final area of one cycle of investigations. The swinging between the two roles should be considered too.

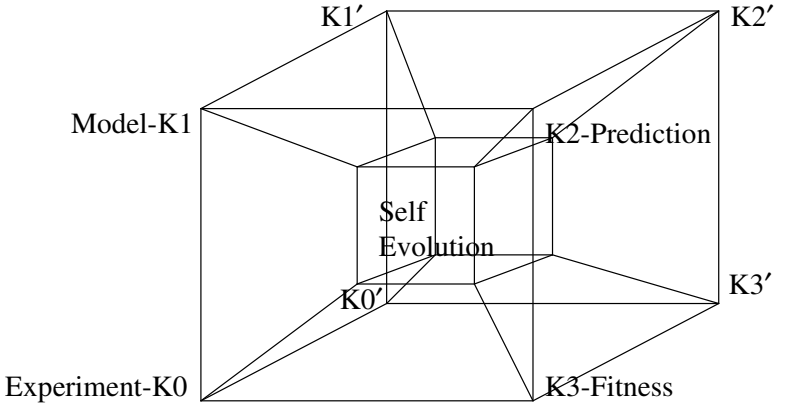

Fig. 11.10 Polytope for self-experimentation system

This suggests that after the integration or direct way we need to look at the differentiation or reverse way. Making use of the developments of the direct way will give different results for the reverse way. The swinging from direct to reverse investigation is beneficial for model evaluation and evolution.

In self-evolvable experimentation, artificial intelligence techniques are employed to carry out the entire cycle of cognition including the elaboration of hypothesis to explain observations, the design of experiments to test these

hypotheses and the physical implementation of the experiments using laboratory automats to falsify hypotheses.

Investigating surprising observations, defined as those observations that disagree with a well-performing hypothesis, has been highlighted as a technique utilized by successful experimenters and has also been considered in previous computational scientific discovery techniques (Lovel et al. 2011).

A surprising observation either highlights a failure in the hypothesis or an erroneous observation. If the observation is highlighting a failure of a hypothesis, especially an otherwise well performing hypothesis with a high prior confidence, then additional experiments should be performed to further investigate the behavior where that observation was found, to allow the development of improved hypotheses. As such we consider the use of surprise to manage the direct way-reverse way trade-off, where obtaining surprising observations will lead to more direct way experiments, and unsurprising observations lead to reverse way developments.

A mathematical formulation for surprise has been considered previously in the literature is Kullback-Leibler divergence used to identify surprising improvements to the models being formed.

The informational distance DD defined by eq. 4.5 plays a similar role.

The DD may be obtained associating similarity matrices to experimental recordings and predictions (Iordache 2009).

Large value of DD states that the observation was surprising, as the overall confidence of the hypotheses has been reduced. A low value of DD states the observation was not surprising, as the overall confidence has increased. The result of DD can therefore be used to control the swinging between direct way and reverse way experiments. A large value for DD will dictate that the next experiment will be integrative, so as to allow investigation of the surprising observation. A low value of DD will lead to a reverse way experiment next, a differential one to search for new surprising features of the behavior.

Thus a hypothesis is removed from consideration if the information loss caused by the removal is small.

After DD has been calculated, the hypothesis manager will go through the process of creating new hypotheses.

This process of evaluating experiments using surprise to choose the next experiment type is automatically continued until the maximum number of experiments allowed has been performed (Lovel et al 2011).

In the coming decades a confluence of wireless networks and lab-on-chip sensor technology with application in health monitoring is expected. In such lab-on-chip network each sensor node is endowed with a limited supply of chemicals. The network will collectively or via the self-evolution level decide how the drug resources will be spent.

Environmental monitoring and improving new drugs and new material discoveries may be performed by similar autonomous experimentation architectures.

References

Ashdown, M.L.: Timed ablation of regulatory CD4 $^{+}$ T cells can prevent murine AIDS progression. J. Immunol. 172, 4917–4925 (2004)

Bailey, R.A.: Efficient semi-Latin squares. Statistica Sinica 2, 413–437 (1992)

Basak, C.S., Zhu, Q., Mills, D.: Prediction of anticancer activiyty of 2-phenylindoles: Comparative molecular field analysis versus ridge regression using mathemathical descriptors. Acta Chim. Slov. 57, 541–550 (2010)

Cioppa, T.M., Lucas, T.W.: Efficient Nearly Orthogonal and Space-filling Latin Hypercubes. Technometrics 49(1), 45–55 (2007)

Coventry, B.J., Ashdown, M.L., Quinn, M.A., Markovic, S.N., Yatomi-Clarke, S.L., Robinson, A.P.: CRP identifies homeostatic immune oscillations in cancer patients: a potential treatment targeting tool? Journal of Translational Medicine 7(102), 1–8 (2009)

Hedayat, A.S., Sloane, N.J.A., Stufken, J.: Orthogonal Arrays. Theory and Applications. Springer, New York (1999)

Iordache, O.: Evolvable Designs of Experiments Applications for Circuits. J. Wiley VCH, Weinheim (2009)

Iordache, O.: Polystochastic Models for Complexity. Springer, Heidelberg (2010)

Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Review 51(3), 455–500 (2009)

Lohr, S.L.: Hasse Diagram in Statistical Consulting and Teaching. The American Statistician 39(4), 376–381 (1995)

Lovell, C., Jones, G., Gunn, S., Zauner, K.-P.: Autonomous Experimentation: Active Learning for Enzyme Response Characterisation. In: JMLR: Workshop and Conference Proceedings, vol. 16, pp. 141–155 (2011)

Lovel, C.J., Zauner, K.-P.: Towards algorithms for autonomous experimentation. In: Naef, F., Galliot, B., Nehaniv, C. (eds.) Proceedings of the 8th International Conference on Information Processing in Cells and Tissues (IPCAT 2009), pp. 150–152 (2009)

Matsumaru, N., Centler, F., Zauner, K.-P., Dittrich, P.: Self-Adaptive Scouting—Autonomous Experimentation for Systems Biology. In: Raidl, G.R., Cagnoni, S., Branke, J., Corne, D.W., Drechsler, R., Jin, Y., Johnson, C.G., Machado, P., Marchiori, E., Rothlauf, F., Smith, G.D., Squillero, G. (eds.) EvoWorkshops 2004. LNCS (LNAI), vol. 3005, pp. 52–62. Springer, Heidelberg (2004)

Nguyen, N.-K.: A new class of orthogonal Latin hypercubes (2008)

Taguchi, G.: Introduction to Quality Engineering: Design Quality into Products and Processes Tokyo. Asian Productivity Organization (1986)

Taguchi, G.: Systems of Experimental Design, vol. 1 & 2. Unipub/Kraus International, White Plains (1987)

Wood, J., Rose, V.S.: Method of designing chemical substances. PCT WO 99/26901 (1999)

Woodcock, J., Woosley, R.: The FDA critical path initiative and its influence on new drug development. Annu. Rev. Med. 59, 1–12 (2008)

Ye, K.Q.: Orthogonal Latin hypercubes and their application in computer experiments. J. Amer. Statist. Assoc. 93, 1430–1439 (1998)