Strategy for finding Chaocipher machine.

I offer the following as a suggested approach. It is derived from part of the Byrne Exhibits provided, but I've no reason to think it deviates when all the data is included.

The first objective of the strategy is to specify unique features expected in the output of a Chaocipher machine, by analysis of the plain and ciphertexts supplied by Byrne.

The second objective is to find a machine, by trial and error, which will meet the required specification. There might be several of these.

The final objective is to discover the keys for the successful machine(s), employing a suitable stochastic algorithm (probably Simulated Annealing or one of its derivatives), using the plaintext to score decrypts of the ciphertext.

Specification of the output from the Chaocipher machine.

This specification is based on observations made by Greg Mellen, Jeff Hill and Moshe Rubin principally and a contribution from myself. It applies to lengthy ciphertexts, of 5000 letters or more.

The data below is derived from analysis of the 5500 letters of plaintext provided in Exhibit 1 of 100 consecutive rows of
ALLGOODQQUICKBROWNFOXESJUMPOVERLAZYDOGTOSAVETHEIRPARTYW

Randomness. Ciphertext must be random in the following respects:

1. when a list of cipher letters is made for each plain letter, all the letters of the alphabet will be in the list;

2. the frequency of each letter of the alphabet in the ciphertext will be nearly equal;

3. when the 5500 plain letters are enciphered with letters chosen at random, the cipher digraphs for plain 'LL' will include anything from 3 to 14 repeats, with an average of 7. The same is true for 'OO' and 'QQ'.

These frequencies also hold for Chaocipher. The frequencies found of repeated cipher digraphs are:

LL = 9; OO = 9; QQ = 10.

4. Analysis of the spacing between repeated cipher digraphs for 'LL', 'OO and 'QQ' in random encipherment provides this pattern:

	spacing, number of rows apart									
from	1	10	20	30	40	50	60	70	80	90
to	9	19	29	39	49	59	69	79	89	99
% of all repeated	18	18	15	13	11	9	7	5	3	1

digraphs

A similar pattern is found in the Chaocipher data.

5. When 5500 letters are created randomly and a search is made for repeated pentagraphs, one can expect to find anything from none to 6 repeats. When this is done many times the average number of repeated pentagraphs is 1.3

Analysis of the 5500 cipher letters of Exhibit 1 finds 4 repeated cipher pentagraphs (PQHMN, LQYMR, DLNAA, MOWLH) which is in line with expectations from random letters. Note that for the complete ciphertext of 13336 letters Moshe Rubin found 10 repeated cipher pentagraphs. When you do the maths, this also fits with what would be expected from encipherment with letters chosen at random.

Non-Randomness.

6. As discovered by Greg Mellen and Jeff Hill, there is evidence of 13-letter blocks in the ciphertext and this should be replicated in any putative Chaocipher machine. Jeff Hill's procedure is ideal. I followed this for the 5500 plain and cipher letters with results as below.

step	frequency of repeated cipher letter	frequency of repeated cipher & plain letter	expected frequency
1	193	0	7
2	240	0	9
3	204	0	7
4	234	0	9
5	227	0	8
6	208	0	7
7	226	0	8
8	198	0	7
9	208	1	8
10	197	0	7
11	233	5	9
12	215	13	8
13	211	0	8
14	224	34	8
15	221	14	8
16	193	5	7
17	220	4	8
18	199	5	7
19	209	1	8
20	192	0	7

7. In these plain/cipher repeats, the plain letter is always 'O' until step 13. Another indication of blocks of 13.

step	letters at	code	plain
9	4207,4216	E,E	O,O
11	389,400	A,A	O,O
11	994,1005	W,W	O,O
11	3689,3700	W,W	O,O
11	3854,3865	N,N	O,O

11	4624,4635	M,M	O,O
12	895,907	W,W	O,O
12	1072,1084	S,S	O,O
12	1347,1359	B,B	O,O
12	1445,1457	E,E	O,O
12	1500,1512	N,N	O,O
12	1940,1952	P,P	O,O
12	2007,2019	A,A	O,O
12	3205,3217	P,P	O,O
12	3480,3492	O,O	O,O
12	4097,4109	E,E	O,O
12	4757,4769	H,H	O,O
12	5350,5362	D,D	O,O
12	5405,5417	I,I	O,O