ELSEVIER

Journal of Molecular Graphics and Modelling 21 (2003) 449–462

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

Assessing the performance of OMEGA with respect to retrieving bioactive conformations

Jonas Boström [a,*], Jeremy R. Greenwood [b], Johan Gottfries [a]

[a] *Department of Medicinal Chemistry, AstraZeneca R&D Mölndal, S-431 83 Mölndal, Sweden*

[b] *Department of Medicinal Chemistry, Royal Danish School of Pharmacy, Universitetsparken 2, DK-2100 Copenhagen, Denmark*

Received 28 June 2002; accepted 14 November 2002

Abstract

OMEGA is a rule-based program which rapidly generates conformational ensembles of small molecules. We have varied the parameters which control the nature of the ensembles generated by OMEGA in a statistical fashion (D-optimal) with the aim of increasing the probability of generating bioactive conformations. Thirty-six drug-like ligands from different ligand–protein complexes determined by high-resolution (≤2.0 Å) X-ray crystallography have been analyzed. Statistically significant models ($Q^2 \geq 0.75$) confirm that one can increase the performance of OMEGA by modifying the parameters. Twenty-eight of the bioactive conformations were retrieved when using a low-energy cut-off (5 kcal/mol), a low RMSD value (0.6 Å) for duplicate removal, and a maximum of 1000 output conformations. All of those that were not retrieved had eight or more rotatable bonds. The duplicate removal parameter was found to have the largest impact on retrieval of bioactive conformations, and the maximum number of conformations also affected the results considerably. The input conformation was found to influence the results largely because certain bond angles can prevent the bioactive conformation from being generated as a low-energy conformation. Pre-optimizing the input structures with MMFF94s improved the results significantly.

We also investigated the performance of OMEGA in connection with database searching. The shape-matching program Rapid Overlay of Chemical Structures (ROCS) was used as search tool. Two multi-conformational databases were built from the MDDR database plus the 36 compounds; one large (maximum 1000 conformations/mol) and one small (maximum 100 conformations/mol). Both databases provided satisfactory results in terms of retrieval. ROCS was able to rank 35 out of 36 X-ray structures among the top 500 hits from the large database.

Keywords: Conformational analysis; Bioactive conformations; OMEGA; High-resolution X-ray crystallography; Shape-searching

1. Introduction

Searching databases with the aim of finding structures that are similar in some fashion to a given query structure is of great interest in ligand-based design. Such searches are typically based on either molecular graph techniques (2D) [1] or superposition (3D) methods [2]. Historically, the 3D methods have been less frequently used, mainly for the reason that they have been much slower than 2D methods. However, the great advantage of using 3D over 2D methods is that they explicitly take shape-dependent properties into account. In this way, 3D methods increase significantly the probability of finding structures which have similar shape and chemical properties to the query structure, but which are less intuitively related according to chemical class.

When using a 3D approach where a conformational database is not required as input (e.g. ISIS [3] and Unity [4]), the query molecules are template-forced to match the single conformations in the database. We consider such approaches less physical; it is preferable to compare ensembles of conformations known in advance to be accessible. A pre-requisite for obtaining reliable results when using pre-calculated conformations (e.g. OMEGA and Catalyst) [5], is that representations of the ligands bioactive conformation must be present in the multi-conformer database.

In a previous comparison study of the ability of various conformational tools to reproduce bioactive conformations [6], we found that the OMEGA program [7] performed competitively. In particular, an exhaustive conformational search on a medium-sized single molecule only takes a fraction of a second with OMEGA, in contrast to other commonly used tools. This previous study, conducted with default parameters, suggested OMEGA as a suitable tool for the generation of databases containing multiple conformations.

* Corresponding author. Tel.: +46-31-706-5251; fax: +46-31-776-3710.
E-mail address: jonas.bostrom@astrazeneca.com (J. Boström).

1093-3263/02/$ – see front matter © 2002 Elsevier Science Inc. All rights reserved.
PII: S1093-3263(02)00204-8

Table 1
The ligand studied, the PDB code, some crystallographic parameters and the protein in the ligand–protein complex

No.	PDB code	Ligand	Resolution (Å)	*R*-factor	*B*-factor[a]	Protein
1	1a28	Progesterone	1.80	0.191	20–32	Progesterone receptor
2	1tng	Aminomethylcyclohexane	1.80	0.172	10–20	Trypsin
3	1tnh	*p*-Fluorobenzylamine	1.80	0.168	10–25	Trypsin
4	1qft	Histamine	1.25	0.184	9–11	Histamine-binding protein
5	1ftm	AMPA	1.70	0.210	7–15	Glutamate receptor-2
6	1phg	Metyrapone	1.60	0.190	12–19	Cytochrome P450-cam
7	3bto	3-Butyl-thiolane-1-oxide	1.66	0.207	12–21	Liver alcohol dehydrogenase
8	1ia3	5-[4-*tert*-Butylphenylsulfanyl]-2,4-quinazolinediamine	1.78	0.160	10–20	Candida albicans dihydrofolate reductase
9	1fcy	6-(5,5,8,8-Tetramethyl)-5,6,7,8-tetrahydro-naphtalene-2-carbonyl)-naphtalene-2-carboxylic acid	1.30	0.134	15–32	Human retinoic acid nuclear receptor
10	1d3g	2-Biphenyl-4-yl-6-fluoro-3-methyl-quinoline-4-carboxylic acid	1.60	0.168	17–24	Human dihydroorotate dehydrogenase
11	1c83	6-(Oxalyl-amino)-1*H*-indole-5-carboxylic acid	1.80	0.197	8–12	Tyrosine phosphatase 1B
12	1ecv	5-Iodo-2-(oxalyl-amino)-benzoic acid	1.95	0.194	7–26	Tyrosine phosphatase 1B
13	1fcz	BMS181156	1.38	0.132	13–24	Retinoic acid nuclear receptor
14	1gr2	Kainate	1.90	0.216	9–14	Glutamate receptor-2
15	1ian	4-[5-(3-Iodo-phenyl)-2-(4-methanesulfinyl-phenyl)-1*H*-imidazole-4-yl]-pyridine	2.00	0.234	13–34	P38 map kinase
16	1frb	3,4-Dihydro-4-oxo-3-((5-trifluoromethyl-2-benzothiazolyl)methyl)-1-phthalazine acetic acid	1.70	0.158	2–13	Fr-1 protein
17	1bju	1-(4-Amidinophenyl)-3-(4-chlorophenyl)urea	1.80	0.171	9–31	Trypsin
18	1dyr	Trimethoprim	1.86	0.181	10–20	Dihydrofolate reductase
19	2izg	Biotin	1.36	0.206	15–23	Streptavidin
20	1cbx	L-Benzylsuccinate	2.00	0.166	5–21	Carboxypeptidase A
21	5std	(6,7-Difluoroquinazolin-4-yl)-(1-methyl-2,2-diphenyl-ethyl)-amine	1.95	0.212	11–23	Scytalone dehydratase
22	6std	2,2-Dichloro-1-methanesulfinyl-3-methyl-cyclopropanecarboxylic acid-[1-(4-bromo-phenyl)-ethyl]-amide	1.80	0.195	18–29	Scytalone dehydratase
23	7std	Carpropamide	1.80	0.197	2–24	Scytalone dehydratase
24	1cbs	Retinoic acid	1.80	n.a.[b]	9–16	Retinoic acid-binding protein
25	1dam	6-(5-Methyl-2-oxo-imidazolidin-4-yl)-hexanoic acid	1.80	0.180	15–25	Dethiobiotin synthetase
26	3std	(3-Aminomethyl-cinnolin-4-yl)-(3,3-diphenyl-allylidene)-amine	1.65	0.194	10–16	Scytalone dehydratase
27	1ejn	*N*-(1-Adamantyl)-*N*′-(4-guanidinobenzyl)urea	1.80	0.200	17–32	Urokinase
28	1if8	(*S*)-*N*-(3-Indol-1-Yl-2-methyl-propyl)-4-sulfamoyl-benzamide	1.94	0.220	12–36	Carbonic anhydrase II
29	1caq	3-(1*H*-Indol-3-yl)-2-[4-(4-phenyl-piperidin-1-yl)-benzenesulfonylamino]-propionic acid	1.80	0.193	12–22	Stromelysin-1
30	1mtv	(+)-2-[4-[(-1-Acetimidoyl-4-piperidinyl)oxy]-3-(7-amidino-2-naphthyl)]-propionic acid	1.90	0.169	10–32	Trypsin
31	1mtw	(+)-2-[4-[((*S*)-1-Acetimidoyl-(3*S*)-pyrrolidinyl)oxy]-3-(7-amidino-2-naphthyl)]-propionic acid	1.90	0.169	6–29	Trypsin
32	1pph	3-TAPAP	1.90	0.167	8–23	Trypsin
33	1f0u	RPR128515	1.90	0.172	11–29	Trypsin
34	1fkg	(1*R*)-1,3-Diphenyl-1-propyl (2*S*)-1-(3,3-dimethyl-1,2-dioxopentyl)-2-piperidinecarboxylate	2.00	0.184	2–17	Fk506-binding protein
35	1fkh	(1*R*)-1-Cyclohexyl-3-phenyl-1-propyl-(2*S*)-1-(3,3-dimethyl-1,2-dioxopentyl)-2-piperidinecarboxylate	1.95	0.161	3–24	Fk506-binding protein
36	1ppc	NAPAP	1.80	0.181	14–31	Trypsin

[a] Interval for the ligand temperature factors.
[b] *R*-value not available.

The aim of the present study was to investigate whether modifying the parameters which control the OMEGA conformation ensemble size can generate ensembles in which there is a higher probability that the bioactive conformation is present. We also examine here the effects of different OMEGA settings on retrieval, in terms of discriminating between different ligands, from two multi-conformational databases. The shape-matching program Rapid Overlay of Chemical Structures (ROCS) [5b,8] was used for this purpose.

2. Selection of ligand–protein complexes

Because the unmodified X-ray structure of the protein-bound ligand is compared to conformations obtained by calculation, the ligand–protein complexes need to be of very high resolution. That is, the resolution of the complexes must be sufficient to unambiguously provide the bioactive conformation. However, in some cases even though the resolution of the ligand–protein complex is high, the atomic positions of the ligand itself may be ill-defined [6]. Thus, only ligands with both high-resolution and low temperature factors (*B*-factors) can be used as test cases. The ligands used in this study were taken from the Brookhaven Protein Data Bank [9]. ReLiBase[1] was used to help select the molecules [10]. The following criteria were used:

- The X-ray structure resolution must be high (≤2.0 Å).
- The *B*-factors of the ligands must be low (preferably below 30).
- The ligands must not include rotatable bonds that cannot be detected by protein crystallography, e.g. hydroxyl torsions.
- The ligands should not include unusual moieties.
- The ligands must be reasonably small, flexible and drug-like.

A total of 36 ligands were found, meeting these criteria. The ligands in the set have 1–11 rotatable bonds, with a roughly even distribution over this range. Data for the ligands are given in Table 1 and the molecular structures are shown in Fig. 1.

3. Computational methods

3.1. OMEGA (version 1.0b4)

OMEGA uses a depth-first searching algorithm for generating conformational ensembles. It is a rule-based method that generates conformations extremely rapidly. OMEGA disassembles the molecule into fragments of (up to five contiguous) rotatable bonds, and reassembles the fragments based on the sorted order of the fragment energies—the depth-first method. OMEGA uses a modified version of the Dreiding force field, which does not include any electrostatic terms. Note that the input bond lengths, bond angles and ring conformations used by OMEGA are not optimized in the assembly step, and there is no explicit term in the conformational energy to account for variations in these geometric parameters. Standard conformations for various ring systems are by default taken from a ring library file. This utility may be switched off. All heavy atoms are superimposed by a least-squares procedure to test for redundant conformations.

OMEGA is controlled by a configuration file where three parameters may affect the generated conformational ensembles: GP_ENERGY_WINDOW which defines an upper bound relative to the global minimum (used to discard high-energy conformations); GP_RMS_CUTOFF which specifies the RMSD value below which two conformations are considered to be the same; and GP_NUM_OUTPUT_CONFS which is the maximum number of output conformations for each input structure. The conformational searches are terminated when the energy cut-off exceeds the defined threshold (GP_ENERGY_WINDOW) or when the maximum numbers of structures have been built. Since the GP_NUM_OUTPUT_CONFS parameter invokes a pseudo-truncation of the search (i.e. when the number of conformations hits this maximum the search is terminated) the parameter will have the greatest influence on molecules with many torsional degrees of freedom.

It is also possible to manipulate the vdW1–4 clash factor (contained in the parameter file, not in the configuration file) which is a standard term in the Dreiding force field. For 1–4 interactions, the vdW radii of the two atoms are scaled down by this factor before the interaction is calculated. The sum of the vdW term plus the torsion term is thus lowered, with the aim of lowering the rotational barrier and reducing sensitivity to the input conformation.

3.2. The least-squares superimposition procedure

All conformations in a given ensemble were superimposed on the corresponding unmodified X-ray structure by a least-squares procedure. Only non-hydrogen atoms were matched. To remove artificial differences caused by symmetry, the Match3D program was used [11]. Match3D re-numbers the structures that are fitted to each other to avoid obtaining large RMS deviations for conformations which differ primarily in the automorphism group. We also used Match3D's feature of comparing the X-ray structure with both the image and the mirror image of the generated structure, to obtain the lowest possible RMS deviation. Geometries of chiral compounds are not inverted in the fitting procedure.

Two conformations are generally considered to be the same unless the least-squares superimposition of the compared atoms finds one or more pairs of equivalent atoms separated by more than 0.3 Å [12]. In this study, a conformation is considered to be reproduced if the RMSD value is less than 0.5 Å, as compared to the unmodified X-ray structure. Fig. 2 shows a least-squares superposition of the X-ray structure

[1] ReLiBase is a copyright of Manfred Hendlich 1994–1999 and Cambridge Crystallographic Data Centre 1999–2000.

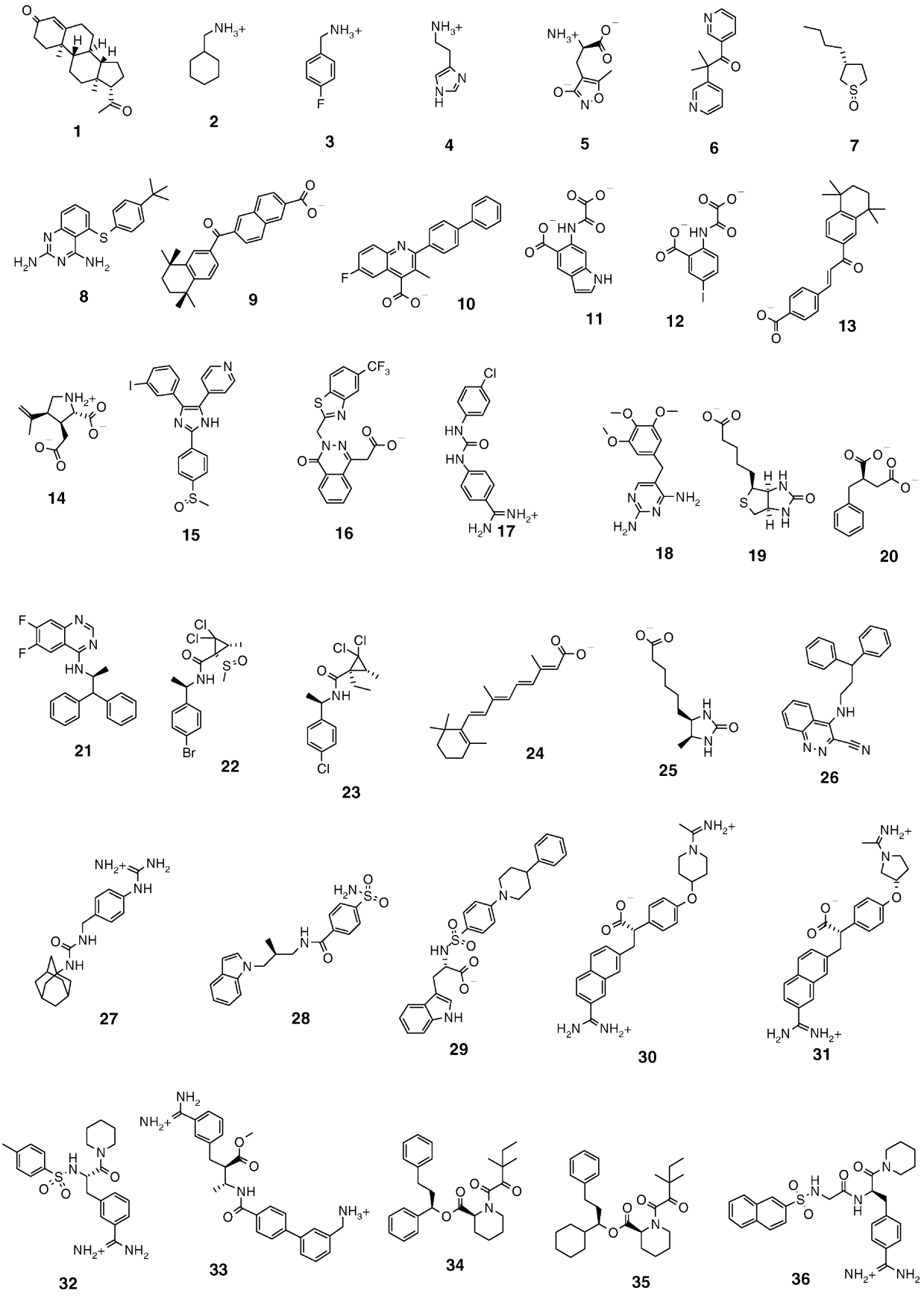

Fig. 1. Structures of the ligands studied in this work.

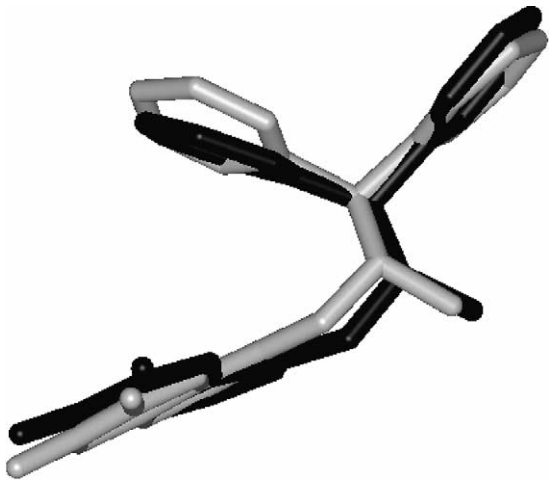

Fig. 2. A least-squares superposition of the unmodified X-ray structure **21** (grey) and the corresponding best conformation (black) in the conformational ensemble generated by OMEGA, as a representative example. The RMSD value is 0.46 Å. The corresponding shape Tanimoto value is 0.92.

21 (grey) and the corresponding best conformation (black) in one conformational ensemble generated by OMEGA, as a representative example. The RMSD value was 0.46 Å.

3.3. Experimental design

OMEGA is controlled by the five main parameters described above. The effects of these parameters were investigated using the following values:

- GP_ENERGY_WINDOW: 3, 5, 8, 10, 20, 40, 100 kcal/mol;
- GP_RMS_CUTOFF: 0.25, 0.4, 0.6, 0.8, 1.0. 1.5 Å;
- GP_NUM_OUTPUT_CONFS: 10, 20, 50, 100, 200, 400, 1000 conformations;
- vdW1–4: 0.6, 0.7, 0.8, 0.95;
- ring library: on, off.

To exhaustively explore all possible combinations above would require 2352 different trials. Therefore, a simplified representative approach was adopted. A subset of 42 out of the 2352 possible setting combinations were selected using D-optimal design using Modde (version 6.0) [13,14] (see Table 2). The objectives were: (i) estimation of the hit-rate, i.e. how many of the possible 36 bioactive conformations were retrieved in the respective searches; and (ii) the average of the RMSD values for the best-fit for each ligand. These two dependent variables (hit-rate and average RMSD) were calculated and applied as regressors in a PLS [15] correlation, using the starting settings as descriptor matrix. Simca (version 8.1) [16] was used for all multivariate correlations. The range for the settings of GP_ENERGY_WINDOW and GP_NUM_OUTPUT_CONFS both exceed one order of magnitude and therefore they were fed to the multivariate calculations in log-scale. The PLS output revealed curvature that was explained by square-terms for GP_NUM_OUTPUT_CONFS and GP_RMS_CUTOFF which are included in all calculations presented under Section 4. Models for each response and each source of input structure were made to find optimal settings for each, resulting in eight different models.

3.4. Different sources of input geometries

OMEGA's dependency on the quality of the input structure was investigated. That is, the 42 settings were tested with four different sources of input geometries, to monitor how the initial conformation influences the results. In the first case, the unmodified X-ray structures were used. In the second case, the structures were generated with the standard 2D-to-3D-generator program Corina [17]. In the third case, SMILES strings were given as input to OMEGA and its own 1D-to-3D SMILES pattern converter was thus used. In the fourth case, the structures obtained from Corina were subjected to rapid geometry optimization (20 iterations) using the MMFF94s force field implemented in MacroModel [18].

3.5. Construction of multi-conformational databases from the MDDR database using OMEGA

The MDL Drug Data Report (MDDR) database [19] contains around 120 000 compounds and is focused on drug-like molecules (i.e. drugs launched or drugs under development), and is therefore a suitable complement for the X-ray structures employed in this study. The MDDR database was converted into a 3D multi-conformational OMEGA databases as follows. The MDDR database was first dumped to a 2D file. This was then pretreated: covalently bonded salts were split; the smallest fragments were removed; hydrogens were added; and charges were added for acids, amines, and amidines. Finally, duplicate entries were removed. The database was subsequently extended with the unmodified X-ray bioactive conformations of ligands **1**–**36**. Corina was then used to convert the 2D structures into 3D representations and to convert the X-ray conformations to Corina structures. All structures were subjected to 20 iterations of geometry optimization (the derivative convergence criterion was set to 0.1 kJ/(Å mol)), using the MMFF94s force field implemented in MacroModel. This last step was performed to obtain appropriate representations of the hard degrees of freedom, such as bond lengths and bond angles, which are not optimized when using OMEGA. The MacroModel "premin" script [18] was used for this purpose. This script identifies problematic structures, i.e. those which are chemically incorrect or have functional groups not covered by the force field parameters, and geometry optimizes the remaining structures. About 90% of the structures were successfully optimized, the rest being retained as Corina structures. The speed is approximately two optimized structures per second on a standard SGI workstation (1 × 300 MHz, R12000).

Two multi-conformational databases were built to investigate the effect of using different parameter settings on retrieving ligands. The OMEGA settings used were based

Table 2
The settings, GP_ENERGY_WINDOW (NRG), GP_NUM_OUTPUT_CONFS (NUM), GP_RMS_CUTOFF (RMS), vdW1–4, ring library (Ringlib), and the results for the 42 different runs, showing the hits, the average RMSD value (Av. RMSD) for the best-fit for each ligand, the average number of conformations (Av. Num) generated and the overall timings

Run	Settings					Source of input structure															
	NRG	NUM	RMS	vdW1–4	Ringlib	X-ray				Corina				MMFF94s				SMILES			
						Hits	Av. RMSD	Av. Num	Time (s)	Hits	Av. RMSD	Av. Num	Time (s)	Hits	Av. RMSD	Av. Num	Time (s)	Hits	Av. RMSD	Av. Num	Time (s)
1	40	100	0.25	0.60	Yes	26	0.54	95	199	24	0.62	90	195	23	0.59	94	197	19	0.69	94	211
2	3	1000	0.25	0.60	Yes	26	0.48	341	87	19	0.63	292	66	24	0.56	323	71	17	0.70	307	74
3	8	10	0.40	0.60	Yes	15	0.73	10	158	14	0.82	10	111	19	0.73	11	114	13	0.92	10	125
4	5	20	0.40	0.60	Yes	21	0.65	19	110	14	0.76	18	63	19	0.70	19	60	12	0.86	19	90
5	100	400	0.40	0.60	no	28	0.45	256	223	25	0.58	244	220	26	0.51	260	232	21	0.65	253	264
6	20	100	0.60	0.60	No	26	0.5	68	179	18	0.65	59	145	24	0.56	66	163	18	0.74	65	176
7	40	200	0.60	0.60	Yes	27	0.49	126	209	23	0.60	114	209	26	0.53	125	219	24	0.61	116	210
8	8	400	0.80	0.60	No	24	0.47	129	161	15	0.67	110	108	22	0.54	125	116	11	0.74	132	131
9	3	20	1.00	0.60	No	16	0.7	13	28	11	0.82	12	35	16	0.72	13	20	9	0.87	12	45
10	100	200	1.00	0.60	No	21	0.51	92	222	14	0.66	81	210	22	0.57	91	217	14	0.71	98	245
11	20	10	1.50	0.60	No	12	0.77	8	175	6	0.89	7	131	12	0.77	8	153	6	0.96	8	161
12	10	50	1.50	0.60	Yes	14	0.68	20	189	7	0.78	19	139	14	0.67	20	146	10	0.79	21	155
13	10	10	0.25	0.70	No	17	0.78	11	149	11	0.85	11	93	18	0.76	11	108	10	0.97	11	111
14	5	200	0.25	0.70	No	28	0.48	125	107	18	0.67	104	74	23	0.56	120	62	12	0.75	108	87
15	40	50	0.40	0.70	No	25	0.56	47	208	20	0.67	44	164	23	0.64	47	193	15	0.77	46	202
16	3	10	0.60	0.70	Yes	15	0.73	10	40	10	0.83	9	21	18	0.76	9	22	10	0.94	9	28
17	100	20	0.60	0.70	Yes	22	0.64	20	203	18	0.72	20	195	21	0.68	20	201	15	0.81	20	209
18	20	1000	0.80	0.70	No	26	0.45	237	249	18	0.61	204	198	23	0.51	223	220	15	0.67	251	219
19	5	50	1.00	0.70	No	19	0.61	25	94	11	0.76	23	63	18	0.65	25	48	9	0.81	24	75
20	10	1000	1.00	0.70	Yes	19	0.55	163	238	13	0.62	137	188	20	0.55	143	198	14	0.65	171	200
21	8	100	1.50	0.70	Yes	14	0.66	29	179	6	0.77	24	121	14	0.65	28	128	10	0.76	30	138
22	3	400	1.50	0.70	Yes	13	0.67	31	58	6	0.83	19	28	13	0.72	17	26	9	0.86	20	33
23	8	20	0.25	0.80	Yes	19	0.68	21	161	14	0.74	20	109	19	0.70	21	111	13	0.88	21	122
24	20	200	0.40	0.80	Yes	27	0.54	136	197	26	0.60	118	188	26	0.54	136	197	23	0.65	128	189
25	10	400	0.60	0.80	No	28	0.47	155	167	17	0.65	146	114	26	0.51	151	128	15	0.74	150	134
26	3	50	0.80	0.80	No	22	0.56	26	25	13	0.75	25	35	20	0.64	25	20	9	0.86	24	40
27	100	100	0.80	0.80	Yes	25	0.56	69	207	22	0.64	68	206	24	0.57	70	216	20	0.64	69	208
28	40	10	1.00	0.80	No	17	0.72	10	187	12	0.80	9	158	19	0.73	10	185	9	0.90	9	185
29	5	1000	1.50	0.80	Yes	14	0.63	37	129	9	0.75	33	80	15	0.64	28	78	9	0.80	36	86
30	100	50	0.25	0.95	Yes	23	0.64	51	200	21	0.69	50	193	20	0.66	51	197	17	0.75	50	199
31	3	100	0.40	0.95	No	22	0.61	53	23	11	0.81	36	22	22	0.64	48	21	9	0.82	41	35
32	8	1000	0.60	0.95	No	26	0.49	278	126	16	0.68	187	97	26	0.51	259	105	13	0.75	240	124
33	5	10	0.80	0.95	Yes	18	0.75	9	51	13	0.84	8	39	15	0.78	9	36	9	0.87	8	66
34	10	200	0.80	0.95	No	22	0.56	77	105	14	0.70	62	76	22	0.57	75	77	11	0.78	68	134
35	20	400	1.00	0.95	Yes	20	0.56	115	212	15	0.65	98	148	21	0.53	109	188	16	0.68	112	209
36	40	20	1.50	0.95	No	13	0.72	12	187	9	0.82	12	138	16	0.72	13	183	6	0.90	14	184
37	100	1000	1.50	0.95	No	13	0.65	54	237	9	0.74	33	202	16	0.64	45	232	7	0.78	59	257
38	8	100	0.60	0.70	Yes	26	0.56	62	162	18	0.66	53	115	26	0.58	59	126	18	0.72	57	139
39	10	100	0.40	0.80	Yes	26	0.55	75	177	24	0.63	66	136	25	0.57	74	153	21	0.70	71	162
40	5	1000	0.60	0.80	No	28	0.44	271	136	15	0.67	183	95	25	0.50	237	93	12	0.76	246	124
41	20	1000	0.80	0.70	Yes	24	0.48	251	248	21	0.55	220	234	24	0.49	246	250	18	0.60	239	270
42	8	20	0.25	0.80	No	18	0.66	21	127	13	0.78	20	76	19	0.70	21	81	10	0.92	21	101

on the results from the parameter optimization described above: an energy cut-off of 5 kcal/mol was specified, the GP_RMS_CUTOFF parameter and the vdW1–4 values were both set to 0.6. The ring library was not used. The two databases differed only in the maximum number of conformations allowed per molecule, the first preferring a small number of conformations per molecule database and the second favoring more conformations. To clarify, the larger allowing up to 1000 conformations per molecule while the smaller being restricted to 100 conformations per molecule. By default, all structures containing 17 or more rotatable bonds were removed. The result was a large database containing 98 824 unique compounds with 38 490 469 conformations and a small database containing the same 98 824 compounds with 5 166 930 conformations.

4. Results and discussion

4.1. Hit-rate as a function of the RMSD cut-off

As mentioned above, a conformation is considered to be reproduced if the RMSD value is less than 0.5 Å, as compared to the X-ray structure. The imposition of the stringent criterion is chosen to reflect our interest in accurately obtaining bioactive conformations, for the purpose of molecular design, pharmacophore elucidation, and shape comparison techniques. For example, we later show that a RMSD greater than 0.5 Å, makes the identification of molecules similar to the bioactive conformation, less likely to be successful. For illustration Fig. 3 shows how the hit-rate varies as a function of the RMSD cut-off. It is immediately apparent that all bioactive conformations are found to an accuracy of 2.0 Å. At 1.0 Å the hit-rate doubles (compared with 0.5 Å) when the input structure to OMEGA is from Corina, whereas importantly there is almost no change in the hit-rate when the pre-optimization step is carried out. Clearly, this graph makes it possible to compare our hit-rate to other studies, in which a different criterion to specify a match to a bioactive conformation is chosen.

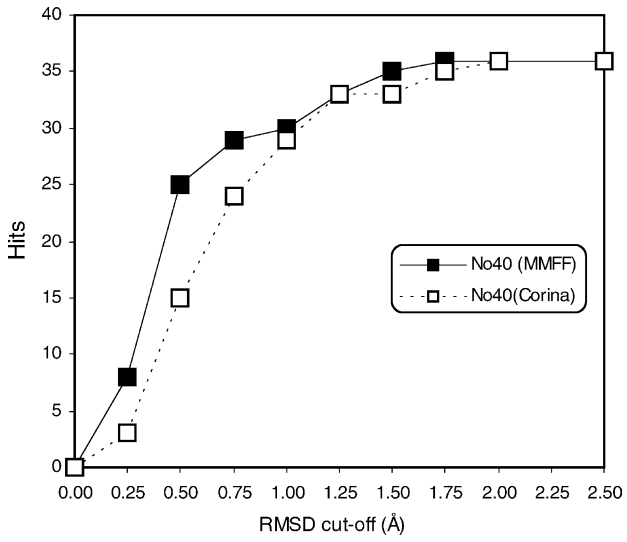

Fig. 3. A graph showing how the hit-rate varies as a function of the RMSD cut-off for one representative setting (Table 2, run no. 40). All bioactive conformations are found to an accuracy of 2.0 Å. At 1.0 Å the hit-rate doubles (compared with 0.5 Å) when the input structure to OMEGA is from Corina, whereas importantly there is almost no change in the hit-rate when the pre-optimization (MMFF) step is carried out.

4.2. Experimental design

Table 2 gives the results for the 42 different runs, showing the hit-rate and the average RMSD value for the best-fit for each ligand. This table gives the results using X-ray, Corina, MMFF94s and SMILES structures as input. The overall timings and the average number of conformations generated are also shown in Table 2.

Statistically significant PLS models of the outcome variables described above were derived from the 42 different OMEGA settings on the set of 36 protein-bound ligand conformations. All models were essentially equally good, expressed in terms of the Q^2-values (Fig. 4). To test for possible chance correlations the dependent variables were shuffled into several random combinations and the models were re-derived. None of the models with scrambled data were found to have any predictive power ($Q^2 < 0.2$), thus ruling out the possibility of chance correlations. The PLS models were used to predict the full candidate set ($n = 2352$) to find and validate potentially optimal settings within and outside the selected 42 settings for each input structure. The most favorable settings were found to be similar for all four methods of generating initial conformations, with the exception of toggling the use of the ring library.

Table 3 gives the results, performed for each input structure, from three additional runs predicted to be among the very best; one “high_nrg”—the best according to the statistics, and two that prioritize a high hit-rate/CPU time ratio. That is, “suggested” which uses a tight energy cut-off and “small” which uses both a tight energy cut-off and small ensemble sizes. Fig. 4 also shows the PLS coefficients for hit-rate (a) and average RMSD (b). The height of each bar indicates its relative importance in the PLS model. For example, the duplicate removal parameter (GP_RMS_CUTOFF) has the largest impact on the hit-rate. These observations are discussed in more detail below.

4.3. The impact of using different input conformations

The input conformation given to OMEGA was found to influence the generated ensembles. The X-ray structure as input performs best. The reason for this is that the search procedure used by OMEGA does not modify the input bond lengths or angles. In the real world bond lengths and angles are dependent on the local chemical environment, such as various substituents and torsional angles. This is reflected

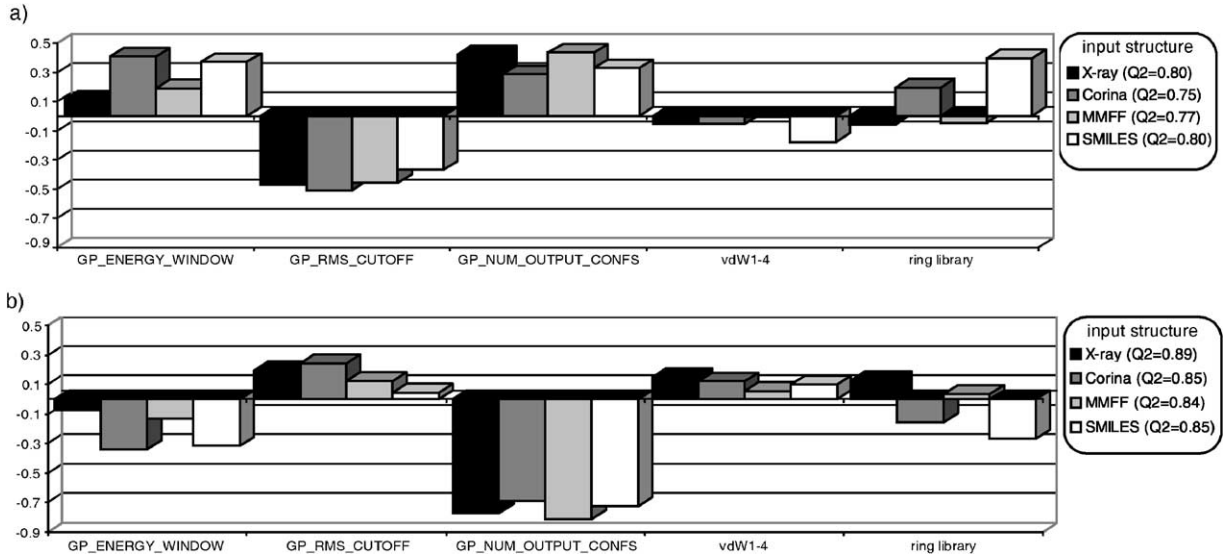

Fig. 4. The PLS coefficients and Q^2-values for the eight different runs with respect to (a) hit-rate and (b) average RMSD. The height of each bar indicates its relative importance in the PLS model.

when using the X-ray structure as input to OMEGA. In general, a greater number of conformations are generated starting from the X-ray structures, because Corina and SMILES exclude certain conformations on the basis of vdW clashes as a consequence of using fixed tables of values for bond lengths and angles. This problem is addressed using a force field such as MMFF94s, which can account to some extent for changes to bond lengths and angles. This is discussed more fully in Section 4.4. For similar reasons it also appears advantageous to use the OMEGA ring library when using Corina structures or SMILES strings as start conformations. That is, more conformations are generated, resulting in a higher hit-rate.

Fig. 5 shows in how many of the 42 runs a ligand is found when starting from the different input structures. The ensembles generated from SMILES strings are the least robust; the number of hits is fewer in most runs. For example, two bioactive conformations (**19**, **24**) that are present in other ensembles are never found in the SMILES (OMEGA) ensembles. By pre-optimizing the Corina structures using MMFF94s, the results were found to be comparable with those obtained using input conformations from X-ray structures. Thus, using a geometry optimization step we see a significant enhancement in hit-rate because the improvement in bond angles relative to experiment enables the search algorithm to correctly retain relevant conformers.

Fig. 5 also shows that molecules with few torsional degrees of freedom are less sensitive to the various settings, whereas the bioactive conformations of structures having eight or more rotatable bonds (**29**–**36**) were not retrieved. To get an idea of how many hits can be possibly achieved, all of the generated conformers were collected, setting the maximum number of conformations to a million and specifying an energy window of 20 kcal/mol, and a very small

Table 3
The results, performed for each input structure, from three runs predicted to be among the very best; one "high_nrg"—the best according to the statistics, and one "suggested" using a tight energy cut-off and one "small" using a tight energy cut-off as well as smaller ensemble sizes

	Source of input structure											
	X-ray			Corina			MMFF94			SMILES		
	high_nrg	Suggested	Small	high_nrg	Suggested	Small	high_nrg	Suggested	Small	high_nrg	Suggested	Small
GP_ENERGY_WINDOW	100	5	5	100	10	10	100	5	5	100	10	10
GP_NUM_OUTPUT_CONFS	1000	1000	100	1000	1000	100	1000	1000	75	1000	1000	100
GP_RMS_CUTOFF	0.6	0.6	0.6	0.4	0.4	0.4	0.6	0.6	0.6	0.4	0.4	0.4
vdW1–4	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
Ringlib	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Hits	28	28	26	27	25	23	26	25	23	26	21	20
Average RMSD	0.43	0.44	0.51	0.53	0.56	0.65	0.49	0.52	0.57	0.52	0.61	0.71
Average number	412	280	44	500	343	55	411	246	42	512	353	58
Time (s)	285	162	98	311	197	144	291	103	51	352	267	160

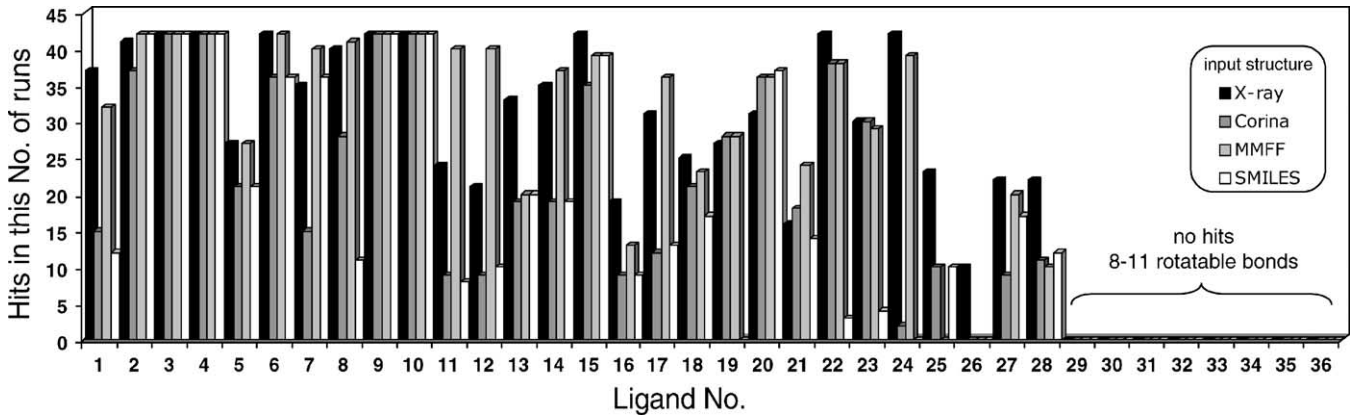

Fig. 5. The number of times a ligand is found when starting from the different input structures. The hit-rate frequency using the X-ray structures as input is greater than when using SMILES strings as input structures. The MMFF94s results were found to be comparable with those obtained using input conformations from X-ray structures. Molecules with few torsional degrees of freedom were less sensitive to the different settings whereas structures having eight or more rotatable bonds were never retrieved.

RMSD cut-off. No additional hits were found. It is not immediately clear why this is the case but two possible causes are apparent: (i) some protein-bound ligands do not bind in a local energy minimum conformation associated with the potential surface defined by OMEGA or (ii) poor parameterization for some fragments.

OMEGA's dependency upon the input structure was not alleviated by altering the force field vdW1–4 parameter. The loadings and PLS coefficients (Fig. 4) for vdW1–4 were close to zero, indicating low impact on the outcome of the dependent variables. Therefore, this parameter may be arbitrarily set anywhere within the examined setting window of 0.6–0.95. In addition, to check whether the hit-rate increases by geometry optimizing the *output* conformational ensembles seeded by Corina, these were post-minimized to their nearest local minimum using MMFF94s/GB/SA. This post-optimization did not improve the results.

4.4. The influence of the GP_ENERGY_WINDOW parameter

The GP_ENERGY_WINDOW parameter defines the energy cut-off which is used to discard high-energy conformations. Unexpectedly, a high-energy cut-off improved the results according to the statistics, particularly when using a rule-based input structure (Fig. 4). This is actually an artifact of the methodology used in the present study. Namely, that there may be other conformations lying within the 0.5 Å RMSD cut-off, which are present in the ensemble and are lower in energy than the best-fit, but which are not recorded because their RMSD values are higher than that of the best-fit. By examining all conformations with an RMSD value less than 0.5 Å, to find the one with the lowest energy, we obtain more reasonable results. Fig. 6 shows the conformational energy penalties for the hits in run no. 24 (Table 2), the most successful run both when seeded by Corina, and when seeded by MMFF94s. Note that structures **24** and **25**, which were found to be hits when starting from one input structure both not the other, has been removed in order to allow comparison.

Using Corina to generate input structures, only one compound (**17**) was found to have a conformational energy penalty higher than 10 kcal/mol, 24 compounds have a conformational energy penalty less than 8.5 kcal, and 20 are found within an energy cut-off of 3.2 kcal (Fig. 6). When using MMFF94s to generate seed structures even better results are obtained. All 25 bioactive conformations show conformational energy penalties within a tolerable energy cut-off (<8 kcal/mol), and all but one (**13**) are found within a satisfactorily low-energy cut-off: 3.2 kcal. Since compound **17** has a reasonably high affinity ($K_i = 16\,\mu M$) for Trypsin [20], such a high conformational energy penalty is very unlikely [21]. The exceptionally high-energy penalty is due to the quality of the input structure. That is, the Corina input structure includes bond angles that prevent the bioactive conformation from being generated as a low-energy conformation. Bond angles θ_1, $\theta_{1'}$ and θ_2 are all 120° in the Corina structure. This set of angles gives rise to an unfavorable vdW clash (H···O=C; see Fig. 7). The corresponding angles in the MMFF94s structure are 128° (θ_1 and $\theta_{1'}$) and 109° (θ_2). The H···O=C distance would then be longer and the vdW clash (and the subsequent high-energy penalty) would thus be avoided (see Fig. 7).

In summary, it appears preferable to use MMFF94s input structures so that a low-energy cut-off can be applied, rather than using a higher energy cut-off and risk sampling exotic conformations. In addition, a low-energy cut-off gives shorter calculation times since the searches are terminated the moment the energy value exceeds the defined threshold.

4.5. The influence of the GP_RMS_CUTOFF parameter

The main purpose of the GP_RMS_CUTOFF parameter is to remove conformations of similar geometry. For example,

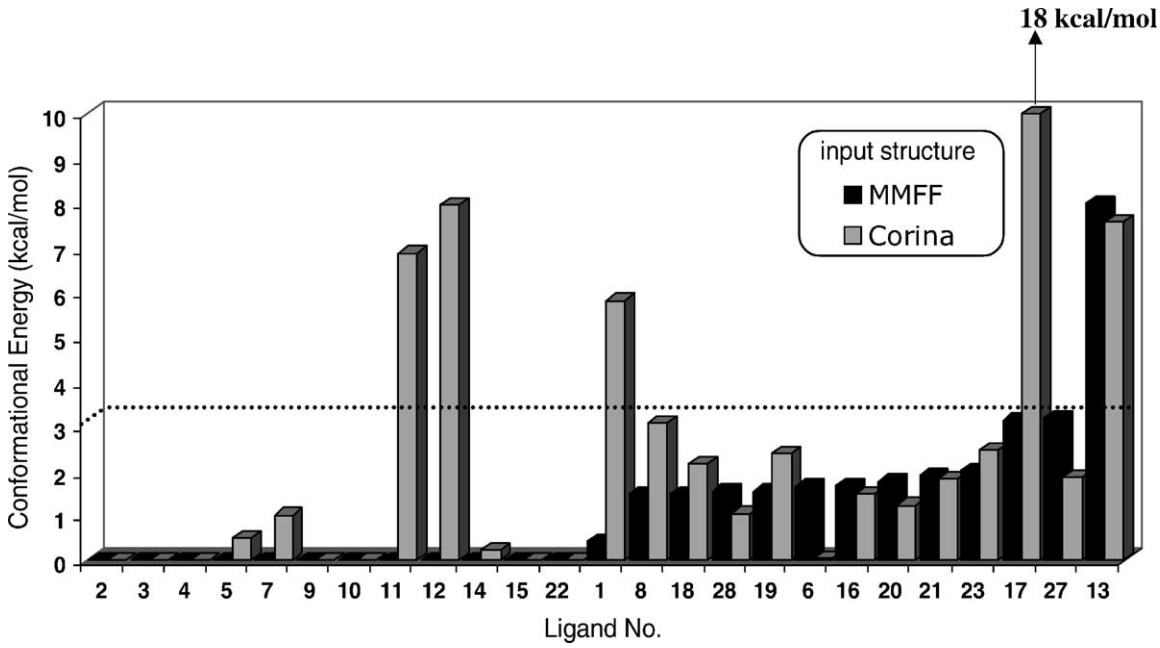

Fig. 6. The conformational energy penalties for the hits in run no. 24, both when seeded by Corina, and when seeded by MMFF94s. In the Corina case, only one compound (**17**) was found to have a conformational energy penalty higher than 10 kcal/mol, 24 compounds have a conformational energy penalty less than 8.5 kcal, and 20 are found within an energy cut-off of 3.2 kcal. MMFF94s performs even better, all 25 bioactive conformations show conformational energy penalties within a reasonable energy cut-off (<8 kcal/mol), and all but one (**13**) are found within a satisfactorily low-energy cut-off: 3.2 kcal.

a higher value can be used to generate diverse conformational ensembles. Fig. 8a shows the response surface for the dependent hit-rate variable against the GP_NUM_OUTPUT_CONFS parameter and the GP_RMS_CUTOFF parameter. Similarly, Fig. 8b plots the average RMSD against the same variables. It can be seen from Fig. 8a that the GP_RMS_CUTOFF parameter has a large impact on the hit-rate. The hit-rate increases progressively by decreasing

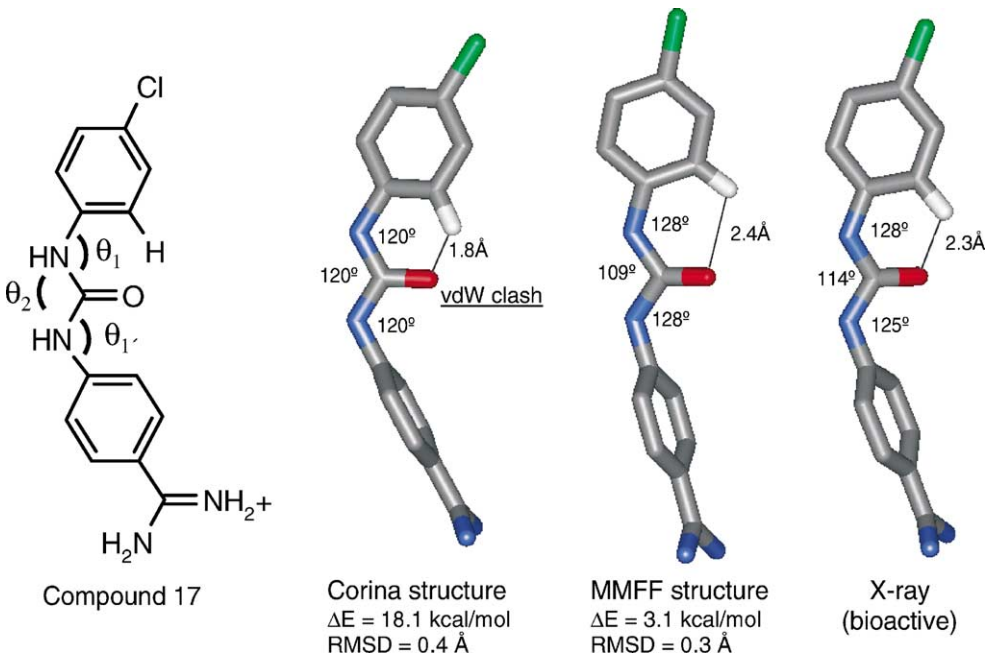

Fig. 7. The reason for the exceptionally high conformational energy penalty for compound **17** may be attributed to the input structure. The Corina input structure includes bond angles that prevent the bioactive conformation from being among the low-energy conformations. That is, the bond angles θ_1, $\theta_{1'}$ and θ_2 are 120° and give rise to unfavorable vdW clashes. The corresponding bond angles in the MMFF94s input structure are 109° (θ_2) and 128° (θ_1 and $\theta_{1'}$) thereby avoiding the vdW clashes and the subsequent high conformational energy penalty.

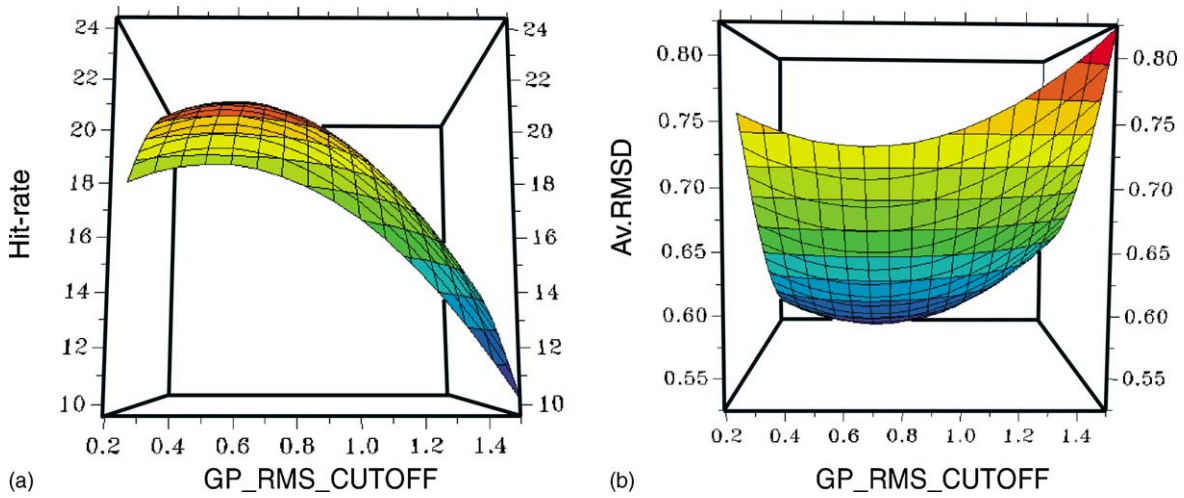

Fig. 8. The response surface for (a) hit-rate variable against the GP_RMS_CUTOFF parameter and the GP_NUM_OUTPUT_CONFS parameter; (b) plots the average RMSD against the same variables. The probability of finding the bioactive conformation reaches its optimum near a GP_RMS_CUTOFF value of 0.6 Å.

the parameter value from 1.5 Å, until it hits a maximum at approximately 0.6 Å, where the graph more or less levels out. For example, run no. 38 versus run no. 21 (Table 2, MMFF94s) clearly indicates that a using low GP_RMS_CUTOFF value (0.6 Å) gives far more hits than using a higher value (1.5 Å)—26 hits versus 14 hits, respectively (Table 2). A parallel effect of the GP_RMS_CUTOFF parameter on the average RMSD value is observed, although this is not as dramatic—0.58 Å versus 0.65 Å, respectively. Thus, the results shown here indicate that it is not an advantage to generate diverse ensembles (with a high value for the GP_RMS_CUTOFF parameter) when attempting to reproduce bioactive ligand conformations. On the contrary, this study indicates that the generation of diverse conformational ensembles will have a negative effect.

4.6. The influence of the GP_NUM_OUTPUT_CONFS parameter

The maximum number of output conformations parameter (GP_NUM_OUTPUT_CONFS) was also found to have a large effect on the dependent variables. Fig. 9a and b shows that the probability of finding the bioactive conformation increases with the number of conformations. This appears obvious since no subset of an ensemble ever truly represents a larger set of conformations. This is particularly true for large, flexible compounds, where the number of low-energy conformations is likely to be very high. A potential solution to this problem is to use a very high value for the GP_NUM_OUTPUT_CONFS parameter, whereby the conformational search is terminated by the energy cut-off

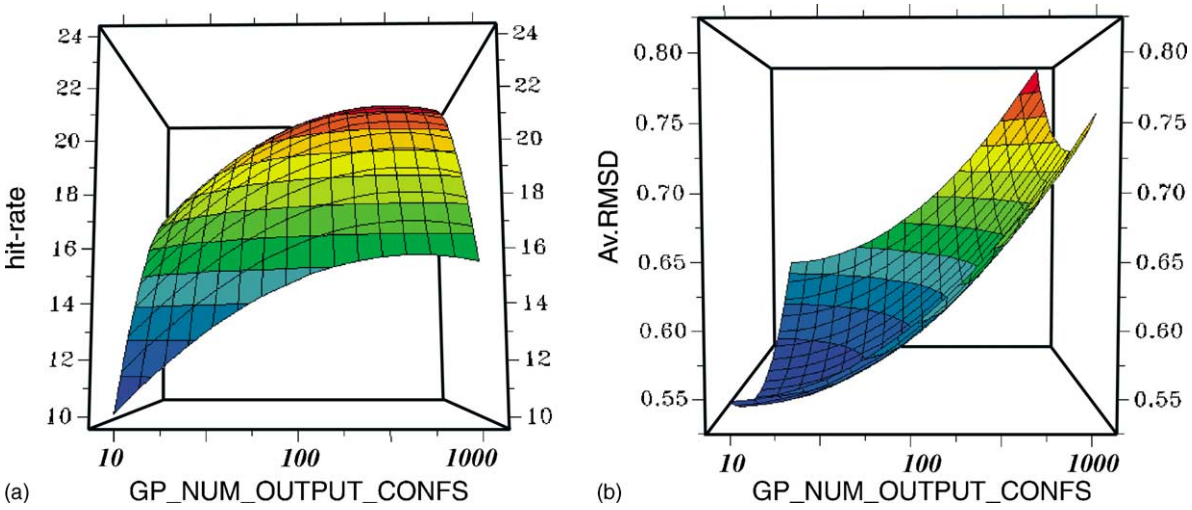

Fig. 9. The response surface for (a) hit-rate variable against the GP_NUM_OUTPUT_CONFS parameter and the GP_RMS_CUTOFF parameter; (b) plots the average RMSD against the same variables. The probability of finding the bioactive conformation increases with the number of conformations in the generated ensembles.

or duplicate removal parameter rather than the maximum number of allowed conformations. However, there may be practical limitations to such an approach, for example increased computational cost and disk-space requirements.

4.7. Database retrieval

An alternative way of defining a hit is by the ability to retrieve the conformation from a database. It is desirable to find out how the suggested OMEGA settings affect the results of searching for a bioactive conformation among the molecules of a database, for example, one containing a hundred thousand structures and millions of conformations. In the ideal case, the bioactive conformation would be present in the OMEGA-generated ensemble, and it would subsequently be ranked top of the list of compounds retrieved as potential hits. However, as we have observed, the bioactive conformation is *not* always present for all ligands, so a perfect match cannot always be made. The main question addressed in this section is: "Even though the conformational ensemble may not contain the ideal bioactive conformation, can the ligand be retrieved within the top 500 most similar compounds, or will the error in the calculated conformations prevent the structures from being correctly identified as a true (shape) match to the query?"

Two multi-conformational databases were built from the MDDR database plus the 36 compounds (see Section 3.5), and then searched. The OMEGA settings used were based on the results from the parameter optimization work described above. Having determined the importance of having a low-energy cut-off and low RMSD value for duplicate removal, the remaining significant variable is the allowed size of the ensemble. Thus, the two databases built differed only in the maximum number of conformations allowed per molecule, one large (termed "suggested" which contained a maximum of 1000 conformations per molecule) and one small (termed "small" which contained maximum 100 conformations/mol).

We employed ROCS for the searching procedure. ROCS is a program that is specifically designed to perform large-scale 3D database searches using a Gaussian shape-based superposition method [5b,8]. Molecules are aligned by a rigid-body optimization process that maximizes the Gaussian overlap volume. The so-called shape Tanimoto is a normalized measure of shape similarity used by ROCS. It is calculated as follows:

$$\text{shape Tanimoto} = \frac{\text{pairwise volume}}{\text{query volume} + \text{hit volume} - \text{pairwise volume}}$$

In order to demonstrate the relationship between the shape Tanimoto and RMSD for this data set, the shape Tanimoto values were calculated for the pairs of conformations that were aligned to minimize the RMSD using Match3D. It should be emphasized that these values of the shape Tanimoto (unlike those used by ROCS), are not optimized for this alignment, but are merely computed for the purpose of comparison. Such shape Tanimoto values are readily computed using the OpenEye Shape-Toolkit [22]. Fig. 10 indicates that a RMSD value of 0.5 Å roughly corresponds to a shape Tanimoto of 0.9. Moreover, Fig. 2 shows a superposition of two conformations of compound **21**. The shape Tanimoto is 0.92 and the corresponding RMSD value is 0.46 Å. Note that this correlation does not seem to hold so well when the compared conformations diverge more from each other. That is, the higher the RMSD the greater the scattering.

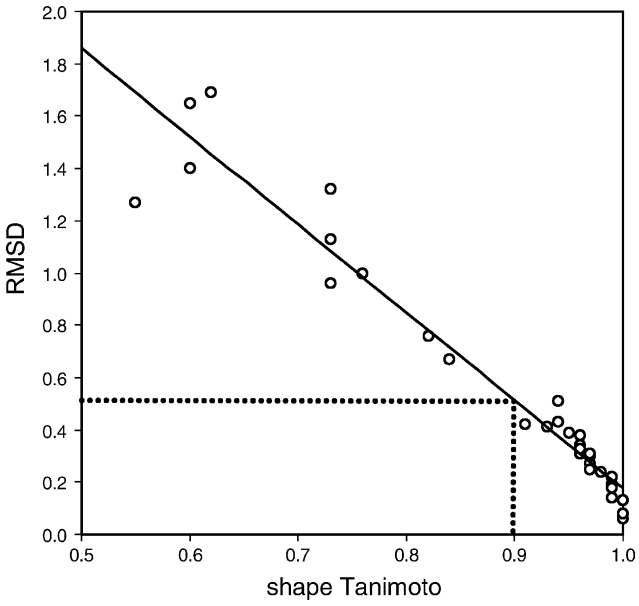

Fig. 10. The RMSD values are plotted against the shape Tanimoto values. The graph indicates that a RMSD value of 0.5 Å roughly corresponds to a shape Tanimoto of 0.9. The correlation does not seem to hold equally well when the compared conformations diverge more from each other. That is, the higher the RMSD, the greater the scattering.

Table 4 gives the results for the database searches, showing the ranking of the query in the hit-list, the shape Tanimoto, and the corresponding RMSD value for the best-fit of each query ligand for both database searches. Both databases provide quite satisfactory results in terms of retrieval. In 28 of the cases the query molecule was ranked among the top 10 molecules when using the "small" database, whereas the query molecules were ranked among the top 10 hits 29 times when using the "suggested" database. The error in the calculated conformations prevented only one structure from being found among the top 500 structures, when searching the "suggested" database. Interestingly, all compounds having eight or more rotatable bonds were found. Four structures (**26**, **33**–**35**) were not found at all when searching the "small" database, and three of these compounds had more than eight torsions. This observation suggests that generating a large number of conformations is important for retrieving the more flexible compounds.

Some cases (**13**, **21**, **28**–**36**) show that even though the conformational ensemble did not contain the ideal bioactive conformation (i.e. shape Tanimoto < 0.9, RMSD > 0.5 Å),

Table 4
The result for the database searches, showing the rank, the shape Tanimoto for the best-fit for each query ligand and the corresponding RMSD value

No.	Rank	Suggested		Rank	Small	
		Shape Tanimoto	RMSD		Shape Tanimoto	RMSD
1	2	0.99	0.70	2	0.99	0.70
2	2	1.00	0.12	2	1.00	0.12
3	1	1.00	0.06	1	1.00	0.06
4	1	0.99	0.26	1	0.99	0.26
5	1	0.96	0.38	1	0.96	0.38
6	2	0.96	1.16	2	0.96	1.16
7	1	0.99	0.19	1	0.99	0.19
8	1	0.97	0.26	1	0.97	0.26
9	1	0.98	0.33	1	0.98	0.33
10	3	0.95	0.37	3	0.95	0.37
11	1	0.94	0.55	1	0.94	0.55
12	1	0.93	0.60	1	0.93	0.60
13	13	0.88	0.76	11	0.88	0.76
14	1	0.99	0.17	1	0.99	0.17
15	1	0.97	0.30	1	0.97	0.30
16	5	0.93	0.51	5	0.93	0.51
17	4	0.94	0.45	4	0.94	0.45
18	1	0.98	0.26	1	0.98	0.26
19	1	0.96	0.41	1	0.96	0.41
20	4	0.94	0.45	4	0.94	0.45
21	37	0.81	2.36	10	0.81	2.36
22	1	0.96	0.37	1	0.96	0.37
23	1	0.95	0.39	1	0.95	0.39
24	2	0.95	0.50	2	0.95	0.50
25	2	0.91	0.56	2	0.91	0.56
26	>500	0.65	2.73	>500	0.61	1.96
27	1	0.92	0.73	1	0.92	0.73
28	1	0.93	0.54	2	0.89	0.65
29	2	0.87	0.78	3	0.84	0.79
30	6	0.82	0.95	61	0.77	1.15
31	88	0.78	1.17	62	0.76	1.37
32	5	0.77	1.36	20	0.71	1.68
33	273	0.75	1.27	>500	0.58	2.46
34	50	0.73	1.72	>500	0.64	1.56
35	150	0.72	4.76	>500	0.66	4.77
36	1	0.87	0.71	1	0.83	0.83

the ligand was found within the top 500 most similar compounds (Table 4). Thus, the shape definition of a hit is more tolerant than when using a RMSD criterion. For this set of compounds the shape Tanimoto had to be above 0.7 for the ligand to be ranked among the top 500. When the bioactive conformation is represented to an accuracy of 0.5 Å, the shape-matching procedure can almost exactly retrieve the ligand from the database (all are ranked within top four).

For the more flexible compounds the shape Tanimoto of a hit was generally found to be higher for the "suggested" database than for the "small", although the difference was not dramatic. The corresponding RMSD values follow the same pattern: 8 of the 36 compounds were found to have better fits, i.e. lower RMSD values for the "suggested" database compared with the "small" database. Consequently, the results indicate a more representative conformational sampling, and thereby better results when using the "suggested" database.

Recalling that ROCS is primarily a shape-based searching system, one should be aware of that a high shape Tanimoto value could be misleading. In some cases, two conformations can adopt very similar overall shapes even though the atom types might differ. For example, the imidazole ring of the histamine in compound **4** may present two very similar shapes for which the atom mappings are completely different. This would lead to a very high shape Tanimoto value, whereas the corresponding RMSD values would differ considerably. This is particularly true for highly symmetric molecules. Nonetheless, despite these limitations, shape-based database querying appears to work surprisingly well.

5. Conclusions

The parameters which control the nature of the conformational ensembles were varied in a statistical manner in order to optimize the performance of OMEGA for the purposes of reproducing bioactive conformations of protein-bound ligands. Thirty-six drug-like ligands determined by high-resolution X-ray crystallography have been analyzed. The data set comprises diverse proteins and ligands it may thus be appropriate to draw general conclusions.

The statistically significant models ($Q^2 \geq 0.75$) reveal that one can increase the performance of OMEGA by modifying the selected parameters. The duplicate removal parameter (GP_RMS_CUTOFF) was found to have the largest impact on the retrieval of bioactive conformations, and the maximum number of conformations (GP_NUM_OUTPUT_CONFS) also affected the results significantly. Molecules with few torsional degrees of freedom were found to be less sensitive to the various settings, whereas the bioactive conformations of structures having eight or more rotatable bonds proved difficult to retrieve. The majority of the structures were found to bind in low-energy conformations, in particular when using MMFF94s to generate input structures. Thus, it appears preferable to pre-optimize the input structures with MMFF94s so that a low-energy cut-off can be used, and thereby avoid sampling exotic conformations. An additional bonus is that a low-energy cut-off gives shorter calculation times. We recommend setting the GP_ENERGY_WINDOW parameter to a low value ($\leq$5 kcal/mol), the GP_RMS_CUTOFF parameter also to a low value ($\leq$0.6 Å), and generating as large conformational ensembles as feasible, with respect to computational cost and available data storage facilities. These settings, in conjunction with ligand pre-optimization, provide optimal performance.

The performance of OMEGA was also investigated in connection with database searching. Two multi-conformational databases were built from the MDDR database plus the 36 compounds; one small and one large, both giving priority to a high hit-rate/CPU time ratio. Thirty-five (97%) hits were found among the 500 most similar compounds in the large database, compared with 32 (89%) in the small

database. Thus, this shape-based database searching appears to work very well. The rank, the shape Tanimoto and the corresponding RMSD values of a hit were in general found to be superior for the large database compared to the small, suggesting that a large number of conformations is important for retrieving highly flexible compounds.

Acknowledgements

The helpful advice from Drs. Andrew Grant, Jens Sadowski, Matthew Stahl and Morten Langgård is gratefully acknowledged.

References

[1] G.M. Downs, P. Willett, Similarity searching in databases of chemical structures, in: K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 7, VCH Publishers, New York, 1995, pp. 1–66.
[2] (a) Y.C. Martin, 3D database searching in drug design, J. Med. Chem. 35 (1992) 2145–2154;
(b) A.C. Good, J.S. Mason, Three-dimensional structure database searches, in: K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 7, VCH Publishers, New York, 1995, pp. 67–117.
[3] T.E. Moock, D.R. Henry, A.G. Ozkabak, M. Alamgir, Conformational searching in ISIS/3D databases, J. Chem. Inf. Comput. Sci. 34 (1994) 184–189.
[4] T. Hurst, Flexible 3D searching: the directed tweak technique, J. Chem. Inf. Comput. Sci. 34 (1994) 190–196.
[5] (a) P.W. Sprague, R. Hoffman, Catalyst pharmacophore models and their utility as queries for searching 3D databases, in: H. Van de Waterbeemd, B. Testa, G. Folkers (Eds.), Computer-Assisted Lead Finding and Optimization, VHCA, Basel, 1990, pp. 230–240;
(b) J.A. Grant, M.A. Gallardo, B.T. Pickup, A fast method of molecular shape comparison. A simple application of a Gaussian description of molecular shape, J. Comp. Chem. 17 (1996) 1653–1666.
[6] J. Boström, Reproducing the conformations of protein-bound ligands: a critical evaluation of several popular conformational searching tools, J. Comput. Aided. Mol. Des. 15 (2001) 1137–1152.
[7] OMEGA (version 1.0b4), OpenEye Science Software, 3600 Cerrillos Road, Suite 1107, Santa Fe, USA, 2001.
[8] ROCS (version 1.0), OpenEye Science Software, 3600 Cerrillos Road, Suite 1107, Santa Fe, USA.
[9] F. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Schimanouchi, M.J. Tasumi, The protein data bank: a computer-based archival file for macromolecular structures, J. Mol. Biol. 112 (1977) 535–542.
[10] M. Hendlich, Acta Crystallogr. D54 (1998) 1178–1182.
[11] Jens Sadowski, AstraZeneca R&D Mölndal, Mölndal, Sweden (personal communication).
[12] J. Sadowski, J. Gasteiger, G. Klebe, Comparison of automatic three-dimensional model builders using 639 X-ray structures, J. Chem. Inf. Comput. Sci. 34 (1994) 1000–1008.
[13] Modde (version 6.0), Umetrics, P.O. Box 7960, Umeå, Sweden.
[14] M.E. Johnson, C.J. Nachtsheim, Some guidelines for constructing exact D-optimal designs on convex design spaces, Technometrics 25 (1983) 271–277.
[15] S. Wold, A. Ruhe, H. Wold, W.J. Dunn III, The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses, SIAM J. Sci. Stat. Comput. 5 (1984) 735–743.
[16] Simca (version 4.5), Umetrics, P.O. Box 7960, Umeå, Sweden.
[17] Corina Molecular Networks, GmbH Computerchemie Langemarckplatz 1, Erlangen, Germany, 2000.
[18] F. Mohamadi, N.G.J. Richards, W.C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrikson, W.C. Still, MacroModel (version 7.1)—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics, J. Comput. Chem. 11 (1990) 440–467.
[19] MDDR—A Structural Database, MDL Information Systems Inc., Prous Science Publishers.
[20] S. Wold, Cross-validatory estimation of the number of components in factor and principal components models, Technometrics 20 (1978) 397–405.
[21] J. Boström, P.-O. Norrby, T. Liljefors, Conformational energy penalties of protein-bound ligands, J. Comput. Aided Mol. Des. 12 (1998) 383–396.
[22] Andrew Grant, AstraZeneca R&D Alderley Park, UK, http://www.eyesopen.com/products/shape.html (personal communication).