UNIVERSITÀ DEGLI STUDI DI TORINO

Medicinal Chemistry Research

Volume 24, Issue 6, 2015

DOI: 10.1007/s00044-015-1333-9

GRID/BIOCUBE4mf to rank the influence of mutations on biological processes to design ad hoc mutants

*Cecilia Rosso, Giuseppe Ermondi, Giulia Caron**

CASSMedChem, DBMSS at the Centre for Innovation, Università di Torino, Via Quarello, 15, 10135 Torino (ITALY)

**email: giulia.caron@unito.it*

ABSTRACT

One of the goals of protein engineering is to design mutants with improved biological profiles, i.e. broader specificity and elevated catalytic activity. Here we propose a novel, fast and general protocol, based on new GRID/BIOCUBE4mf descriptors, to rank mutants for their ability to affect the pattern of interaction with the ligand, and thus their biological profile. The efficacy of the strategy is proven by establishing relationships between a new descriptor (SumΔn%) and Michaelis constants (K_M) for a series of pentalenene synthase mutants.

INTRODUCTION

Since a large number of mutant proteins with single amino acid substitutions are now being produced, the ability to predict the structural changes expected to derive from these mutations would be of great help in guiding mutagenic studies, and in the subsequent interpretation of those changes in terms of stability and mutant-ligand interaction (Zhang, 2008; Chiappori et al., 2009).

Modeling point mutation is a routine task in most molecular-modeling packages. Briefly, two strategies are available: traditional homology modeling, and methods based on rotamer libraries (Dunbrack, 2002; Faber and Matthews, 1990; Eyal et al., 2003; Vasquez, 1996); the latter are faster, since they explore the conformational space of side-chains on a fixed backbone.

The nature of mutant-ligand interactions is generally predicted using docking methods, (Lengauer and Rarey, 1996) which suffer from a number of drawbacks, amply discussed in the literature (Jain, 2009). For example, although the increasing number of reports about new and efficient scoring functions (Spyrakis et al., 2007), metal-ligand interaction still remains a considerable challenge for most computational docking methods (Seeback et al., 2008).

This paper describes a general and fast strategy to predict and compare interactions of mutants with substrates/ligands, the goal being to shed light on changes introduced by any mutation in a given wild-type (WT) structure.

Briefly, mutants obtained from the rotamer library approach are submitted to GRID-BIOCUBE4mf, and characterized in terms of pharmacophoric/physicochemical properties (Sciabola et al., 2010; Caron et al., 2009) in limited subregions, i.e. the regions close to the mutation, by GRID MIFs (see Methods). The method provides a considerable gain in CPU resources compared to traditional homology modeling/docking procedure, and analyses the biosystem in greater depth. For these reasons, it is expected to be of potential interest in the drug discovery field.

To validate the strategy we selected pentalenene synthase, a typical sesquiterpene synthase, of remarkable interest in many research fields (e.g. pharmaceutical and agrochemical) (Seemann et al., 2002), for its potential in the production of terpenoids.

METHODS

3D structures

The crystal structures of pentalenene synthase from *Streptomyces* UC5319 (pdb code: 1PS1) and its mutant N219L (pdb code: 1HM7) were used as reference structures.

The ten mutants (W308F, H309A, H309S, F77Y, H309C, H309F, D84E, D81E, D80E, N219D) were built using the rotamer library tool implemented in MOE 2010.10 (http://www.chemcomp.com) which performs a systematic conformational search of side-chains, followed by the application of an environment-based scoring function to rank the conformers (Soss, 2003). With this method, all sterically reasonable conformations are represented, and it is the particular environment of the side-chain that determines its conformation (rather than *a priori* probabilities based upon the backbone alone).

Protein mobility

To check protein mobility, we used the experimental B-value (or temperature factor) and a normal mode analysis (NMA) for the investigated crystallographic structure (1PS1). Briefly, B values below 10 indicate that the atom maintains about the same position in all molecules in the crystal. B-values above 50 indicate that the atom is moving so fast that it can barely be seen (http://www.rcsb.org/pdb/101/static101.do?p=education_discussion/Looking-at-Structures/coordinates.html). An anisotropic network model (ANM) was used to perform the NMA. Shortly, ANM (Atilgan et al., 2001) is a fast computational tool that provides a simulation of spatial fluctuations of proteins, in good agreement with the CPU time-consuming traditional normal-mode analysis, based on the force fields of molecular dynamics (MD) packages.

GRID/BIOCUBE4mf

We studied the differences introduced by the mutations into the reference structure, employing the GRID/BIOCUBE4mf strategy (Caron et al., 2009). In particular, the effects of the substitution of a single amino acid can be described in terms of patterns of interaction by analyzing simple numerical indices.

The GRID/BIOCUBE4mf approach takes into account an enzyme subregion containing the active site, and combines GRID (Braiuca et al., 2004; Goodford, 1985) with BIOCUBE4mf (Caron et al., 2009). GRID (http://www.moldiscovery.com) is a computational procedure that yields a property distribution map of attractive and repulsive forces between a probe and the target molecule, known as Molecular Interaction Fields (MIFs) (Goodford, 1985; Wade et al., 1993; Wade and Goodford, 1993).

MIFs can draw an accurate picture of the enzyme subregion involved in the substrate/ligand binding. In particular, the GRID hydrophobic ('DRY'), methyl ('C3'), neutral flat amide ('N1') and sp^2 carbonyl oxygen ('O') probes are used to describe respectively hydrophobic, shape, hydrogen bond acceptor and hydrogen bond donor interactions. BIOCUBE4mf filters and computes MIF points with thresholds calibrated in a previous study (Caron et al., 2009) (Table 1).

Table 1. GRID probes and BIOCUBE4mf selected thresholds. The type of interaction between the considered probe and the enzyme is also reported.

Symbol	Probe name	Type of interaction	Threshold [kcal/mol]
DRY	Hydrophobic Probe	Hydrophobic	-0.00005
N1	Neutral flat amide	Hydrogen bond	-8.5
O	sp2 carbonyl oxygen	Hydrogen bond	-7.0
C3	Methyl group	Steric hindrance	-3.3

A GRID box was positioned including the most important residue atoms; the Number of Planes of grid points per Angstrom (NPLA) was set to 2, to give a resolution grid of 0.5 Å. To center the box and include the same residues from any active site, some atoms belonging to the active site of the enzyme were taken as starting point, and their Cartesian coordinates were averaged to obtain the center of the box. In particular, for pentalenene synthase we chose “CA” atoms from Phe76, Phe77, Asp80, Asp81, Asp84, Leu219, Lys227, Arg230, Trp308 and His309. Then, starting from these points, the box was expanded along the three axes, to include all the residues of the active site.

GRID-BIOCUBE4mf results are generally expressed as the number of points below the threshold. In this study, two new descriptors were defined, $\Delta n\%$ and $Sum\Delta n\%$.

Δn% is defined as the difference in number of GRID points below the selected threshold (Table 1) between the mutant and the reference structure (RS, generally the WT), normalized over the RS points and expressed as a percentage (Eq. 2):

$$\Delta n\% = \frac{n_{conf} - n_{RS}}{n_{RS}} * 100 \qquad \text{Eq 2}$$

where n_{conf} is the number of points of the mutant, and n_{RS} is the number of points of the RS (Table 2, last line). According to this definition, Δn% describes the difference between the RS and the investigated mutant for a given probe.

SumΔn% is defined as the sum of the averaged $\Delta n\%_i$ values, where the index i can include all the probes or only some of them (Eq. 3).

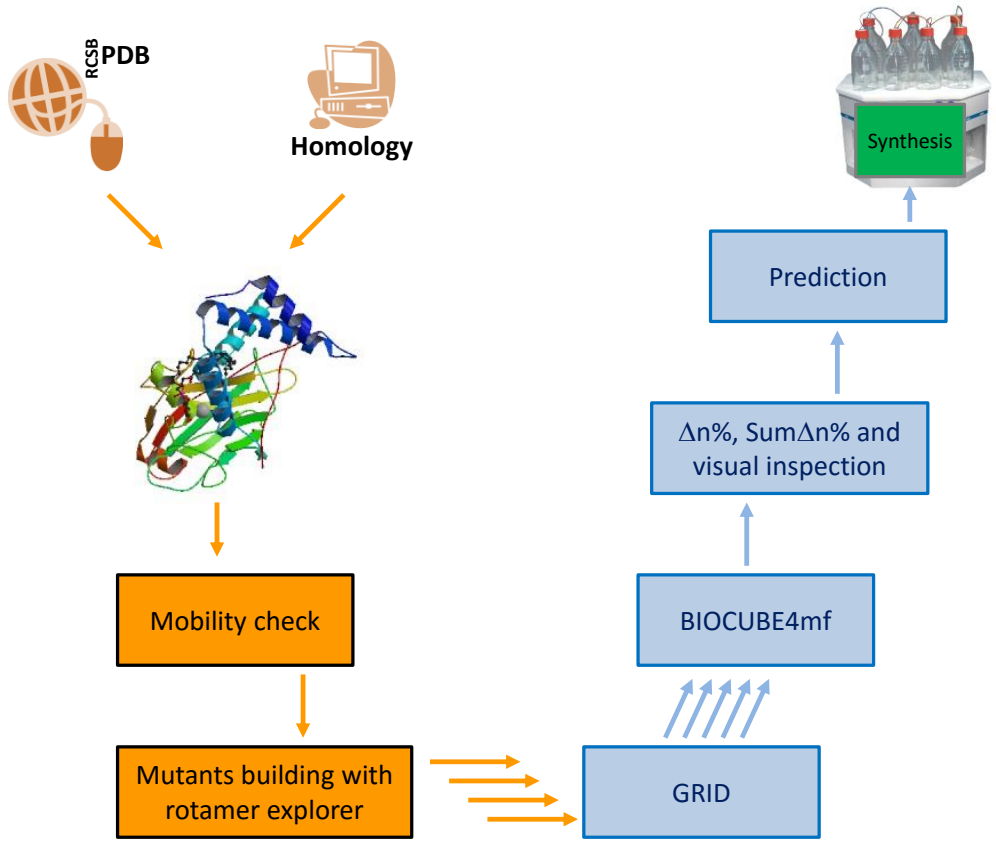
$$Sum\Delta n\% = \sum_i \Delta n\%_i \qquad \text{Eq 3}$$

RESULTS AND DISCUSSION

The setting-up of the adopted protocol

The strategy is schematized by the flow-chart reported in Figure 1.

Figure 1. Flow chart of the novel strategy described in the paper. In orange steps performed using traditional molecular modeling tools (e.g. MOE), in light blue steps performed using the GRID/BIOCUBE4mf approach.



The first step (top left) entails preparing the RS, often the enzyme WT. This latter could be either retrieved from PDB or obtained using standard homology methods, and may thus be of experimental or computational nature.

Once available the 3D structure of the RS, it is mandatory to get information about the binding site since our method does not include the use of substrate/ligand molecules. In practice we have to know which are the residues governing the interaction with the substrate/ligand. This could be obtained with experimental data (i.e. binding studies on some selected examples as in the case of pentalenene synthase, crystallographic data including the substrate/ligand). The presence of this information is the most stringent condition of applicability of the strategy.

The protein collective motions involve large regions of the protein and can influence cavity dimensions and exposure. The second step of the strategy thus consists in verifying the mobility of the binding site. Although in the presence of a rigid binding site, the strategy has a greater possibility of making successfully predictions, in principle there are no limitations in the application of the method to

flexible binding sites. The mobility check is done by B factor analysis and normal mode analysis (NMA) (see Methods).

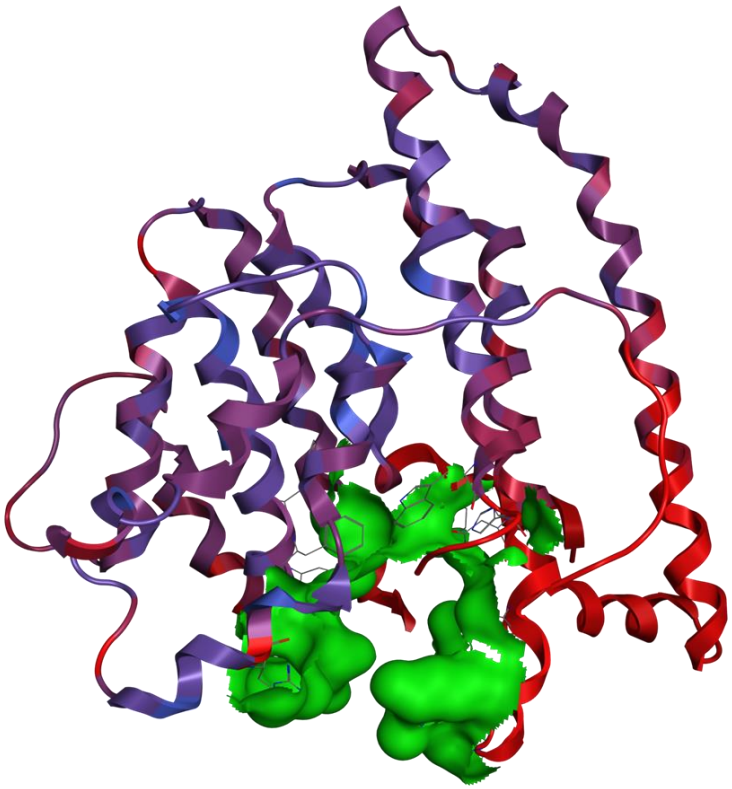
Rotamer explorer is then used to build mutants (third step). For each mutant, the rotamer explorer procedure produces a set of conformers; these are submitted to the GRID/BIOCUBE4mf procedure (fourth step). Finally, quantitative scores and graphical plots are obtained to predict the activity of a set of mutants, and thus to address the production of a narrowed number of selected variants for dedicated catalysis.

Application: Mutants of pentalenene synthase

The strategy illustrated in Fig. 1 was applied to study a series of mutants of pentalenene synthase. from *Streptomyces* UC5319.

The analysis of experimental B factors of the crystal structure (1PS1, Figure 2) revealed that the active site is located in a rather flexible portion of the molecule. Similar results were obtained with the ANM approach (data not shown). This feature makes pentalenene synthase a very interesting and general topic of investigation since many enzymes identified as targets in drug discovery programs have the binding site located in flexible regions.

Figure 2. Crystal structure of pentalenene synthase (1PS1) colored according to experimental B factors (blue = low mobility; red = high mobility), in green the binding site region. The figure shows the presence of a flexible region in the proximity of the binding site. This evidence should be taken into account in the interpretation of the results, see text for details.



The WT and the mutants were then submitted to GRID and subsequently to BIOCUBE4mf, to extract the most relevant information. For the topic under investigation, four GRID probes (DRY, N1, O, C3) were chosen to cover all relevant physicochemical interactions (hydrophobic, electrostatic, and hydrogen bonding) that can be established between the substrate and the functional groups of the active site. Taken together, the analysis of GRID/BIOCUBE4mf data is expected to produce a comparison between the interaction pattern of a given mutant and the analogue profile of the WT. The entire list of GRID-BIOCUBE4mf results is in Table 2. Since the use of default thresholds (see above) for N1 and O probes produced a small number of points (= poor relevance), their contribution was not considered.

Table 2. Numerical outputs of the GRID-BIOCUBE4mf procedure. Each line of the table shows the results for a given rotamer (e.g. F77Y_01). In particular, for any probe (i.e. DRY, C3, N1 and O) we report the number of selected points (i.e. the points in the grid with energy lower than default thresholds reported in Table 1) and Δn% calculated according to Eq. 2. The data of the reference structure (named RS(WT)) used to calculate Δn% are also reported in the last line.

CONFORMER	DRY	Δn%	C3	Δn%	N1	Δn%	O	Δn%
F77Y_01	476	-6.30	215	1.90	4	0.00	4	0.00
F77Y_02	476	-6.30	210	-0.47	4	0.00	4	0.00
F77Y_03	689	35.63	143	-32.23	5	25.00	4	0.00
F77Y_04	696	37.01	122	-42.18	4	0.00	4	0.00
F77Y_05	578	13.78	181	-14.22	5	25.00	4	0.00
F77Y_06	526	3.54	204	-3.32	4	0.00	4	0.00
F77Y_07	478	-5.91	210	-0.47	4	0.00	4	0.00
F77Y_08	664	30.71	116	-45.02	4	0.00	4	0.00
F77Y_09	638	25.59	116	-45.02	4	0.00	4	0.00
F77Y_10	503	-0.98	195	-7.58	4	0.00	4	0.00
F77Y_11	640	25.98	143	-32.23	4	0.00	4	0.00
N219D_01	508	0.00	210	-0.47	3	-25.00	4	0.00
N219D_02	497	-2.17	214	1.42	3	-25.00	4	0.00
N219D_03	486	-4.33	213	0.95	4	0.00	4	0.00
N219D_04	497	-2.17	210	-0.47	3	-25.00	4	0.00
N219D_05	584	14.96	204	-3.32	3	-25.00	4	0.00
N219D_06	542	6.69	204	-3.32	3	-25.00	4	0.00
D80E_10	542	6.69	203	-3.79	4	0.00	4	0.00
D80E_11	469	-7.68	206	-2.37	4	0.00	4	0.00
D80E_12	553	8.86	205	-2.84	4	0.00	4	0.00
D80E_13	468	-7.87	164	-22.27	4	0.00	4	0.00
D80E_14	568	11.81	193	-8.53	4	0.00	4	0.00
D80E_15	530	4.33	152	-27.96	4	0.00	4	0.00
D80E_16	540	6.30	198	-6.16	4	0.00	4	0.00
D80E_17	623	22.64	199	-5.69	4	0.00	4	0.00
D80E_18	478	-5.91	198	-6.16	4	0.00	4	0.00
D80E_19	619	21.85	209	-0.95	4	0.00	4	0.00
D80E_1	451	-11.22	209	-0.95	4	0.00	4	0.00
D80E_20	478	-5.91	176	-16.59	4	0.00	4	0.00
D80E_21	458	-9.84	158	-25.12	4	0.00	4	0.00
D80E_2	525	3.35	190	-9.95	4	0.00	4	0.00
D80E_3	512	0.79	173	-18.01	4	0.00	4	0.00
D80E_4	467	-8.07	236	11.85	4	0.00	4	0.00
D80E_5	459	-9.65	164	-22.27	4	0.00	4	0.00
D80E_6	435	-14.37	210	-0.47	4	0.00	4	0.00
D80E_7	587	15.55	209	-0.95	4	0.00	4	0.00

D80E_8	467	-8.07	201	-4.74	4	0.00	4	0.00
D80E_9	453	-10.83	245	16.11	4	0.00	4	0.00
D81E_10	510	0.39	203	-3.79	4	0.00	4	0.00
D81E_11	514	1.18	190	-9.95	4	0.00	4	0.00
D81E_12	516	1.57	194	-8.06	4	0.00	4	0.00
D81E_13	502	-1.18	192	-9.00	4	0.00	4	0.00
D81E_14	520	2.36	176	-16.59	4	0.00	4	0.00
D81E_15	516	1.57	177	-16.11	4	0.00	4	0.00
D81E_16	503	-0.98	182	-13.74	4	0.00	4	0.00
D81E_17	519	2.17	216	2.37	4	0.00	4	0.00
D81E_18	513	0.98	195	-7.58	4	0.00	4	0.00
D81E_19	514	1.18	203	-3.79	4	0.00	4	0.00
D81E_1	499	-1.77	211	0.00	4	0.00	4	0.00
D81E_20	541	6.50	183	-13.27	4	0.00	4	0.00
D81E_21	548	7.87	182	-13.74	4	0.00	4	0.00
D81E_22	534	5.12	211	0.00	4	0.00	4	0.00
D81E_23	513	0.98	208	-1.42	4	0.00	4	0.00
D81E_24	493	-2.95	219	3.79	4	0.00	4	0.00
D81E_25	496	-2.36	201	-4.74	4	0.00	4	0.00
D81E_26	513	0.98	215	1.90	4	0.00	4	0.00
D81E_27	513	0.98	198	-6.16	4	0.00	4	0.00
D81E_28	521	2.56	205	-2.84	4	0.00	4	0.00
D81E_29	525	3.35	187	-11.37	4	0.00	4	0.00
D81E_2	513	0.98	204	-3.32	4	0.00	4	0.00
D81E_30	521	2.56	192	-9.00	4	0.00	4	0.00
D81E_31	534	5.12	195	-7.58	4	0.00	4	0.00
D81E_32	500	-1.57	235	11.37	4	0.00	4	0.00
D81E_33	503	-0.98	220	4.27	4	0.00	4	0.00
D81E_34	499	-1.77	222	5.21	4	0.00	4	0.00
D81E_35	533	4.92	227	7.58	4	0.00	4	0.00
D81E_36	514	1.18	188	-10.90	4	0.00	4	0.00
D81E_37	583	14.76	180	-14.69	4	0.00	4	0.00
D81E_38	533	4.92	201	-4.74	4	0.00	4	0.00
D81E_39	527	3.74	224	6.16	4	0.00	4	0.00
D81E_3	515	1.38	181	-14.22	4	0.00	4	0.00
D81E_40	516	1.57	236	11.85	4	0.00	4	0.00
D81E_41	519	2.17	197	-6.64	4	0.00	4	0.00
D81E_4	511	0.59	203	-3.79	4	0.00	4	0.00
D81E_5	500	-1.57	201	-4.74	4	0.00	4	0.00
D81E_6	515	1.38	179	-15.17	4	0.00	4	0.00
D81E_7	502	-1.18	205	-2.84	4	0.00	4	0.00
D81E_8	502	-1.18	186	-11.85	4	0.00	4	0.00
D81E_9	507	-0.20	192	-9.00	4	0.00	4	0.00
D84E_10	456	-10.24	225	6.64	4	0.00	4	0.00
D84E_11	456	-10.24	219	3.79	4	0.00	4	0.00
D84E_12	471	-7.28	220	4.27	4	0.00	4	0.00
D84E_13	511	0.59	216	2.37	4	0.00	4	0.00

D84E_14	481	-5.31	222	5.21	4	0.00	4	0.00
D84E_15	455	-10.43	215	1.90	4	0.00	4	0.00
D84E_16	457	-10.04	234	10.90	4	0.00	4	0.00
D84E_17	455	-10.43	214	1.42	4	0.00	4	0.00
D84E_18	567	11.61	211	0.00	4	0.00	4	0.00
D84E_19	502	-1.18	215	1.90	4	0.00	4	0.00
D84E_1	473	-6.89	209	-0.95	4	0.00	4	0.00
D84E_20	490	-3.54	214	1.42	4	0.00	4	0.00
D84E_21	466	-8.27	213	0.95	4	0.00	4	0.00
D84E_22	478	-5.91	218	3.32	4	0.00	4	0.00
D84E_23	491	-3.35	209	-0.95	4	0.00	4	0.00
D84E_24	468	-7.87	215	1.90	4	0.00	4	0.00
D84E_25	478	-5.91	220	4.27	4	0.00	4	0.00
D84E_26	458	-9.84	217	2.84	4	0.00	4	0.00
D84E_27	460	-9.45	212	0.47	4	0.00	4	0.00
D84E_28	479	-5.71	213	0.95	4	0.00	4	0.00
D84E_29	465	-8.46	219	3.79	4	0.00	4	0.00
D84E_2	499	-1.77	213	0.95	4	0.00	4	0.00
D84E_30	472	-7.09	207	-1.90	4	0.00	4	0.00
D84E_31	529	4.13	207	-1.90	4	0.00	4	0.00
D84E_32	506	-0.39	218	3.32	4	0.00	4	0.00
D84E_33	461	-9.25	215	1.90	4	0.00	4	0.00
D84E_34	538	5.91	210	-0.47	4	0.00	4	0.00
D84E_35	480	-5.51	207	-1.90	4	0.00	4	0.00
D84E_36	482	-5.12	217	2.84	4	0.00	4	0.00
D84E_37	533	4.92	207	-1.90	4	0.00	4	0.00
D84E_38	486	-4.33	211	0.00	4	0.00	4	0.00
D84E_39	517	1.77	207	-1.90	4	0.00	4	0.00
D84E_3	462	-9.06	215	1.90	4	0.00	4	0.00
D84E_4	480	-5.51	209	-0.95	4	0.00	4	0.00
D84E_5	533	4.92	216	2.37	4	0.00	4	0.00
D84E_6	482	-5.12	224	6.16	4	0.00	4	0.00
D84E_7	458	-9.84	212	0.47	4	0.00	4	0.00
D84E_8	463	-8.86	214	1.42	4	0.00	4	0.00
D84E_9	473	-6.89	223	5.69	4	0.00	4	0.00
H309A	534	5.12	201	-4.74	3	-25.00	4	0.00
H309C_1	532	4.72	203	-3.79	3	-25.00	4	0.00
H309C_2	534	5.12	221	4.74	5	25.00	6	50.00
H309C_3	534	5.12	204	-3.32	3	-25.00	4	0.00
H309F_1	552	8.66	210	-0.47	5	25.00	4	0.00
H309F_2	542	6.69	208	-1.42	3	-25.00	5	25.00
H309F_3	552	8.66	202	-4.27	3	-25.00	4	0.00
H309F_4	535	5.31	204	-3.32	5	25.00	5	25.00
H309S_1	534	5.12	201	-4.74	3	-25.00	4	0.00
H309S_2	532	4.72	199	-5.69	3	-25.00	4	0.00
H309S_3	532	4.72	207	-1.90	4	0.00	8	100.00

W308F_1	514	1.18	209	-0.95	4	0.00	6	50.00
W308F_2	532	4.72	209	-0.95	4	0.00	4	0.00
RS (WT)	**508**		**211**		**4**		**4**	

Table 2 highlights how the flexibility of a side-chain may modulate the mutant interaction profile, and shows the sensitivity of GRID/BIOCUBE4mf to point up these differences.

At the beginning of the study and after a number of tests, it appeared appropriate to use average values of Δn% as final descriptors to take into account the effect of side-chain flexibility. %Δn profiles are highly informative, but they are also rather complex to use for comparative purposes. We thus sought a numerical score to facilitate the interpretation of the results. We thus defined the sum of %Δn values, including all probes' contribution: Sum%Δn (see Methods for definition). Sum%Δn were then calculated for the ten mutants and discussed below in relation to the available biological data which are listed in Table 3.

The Michaelis constant (K_M) measures the concentration of substrate required for significant catalysis to take place. Although K_M is function of numerous microscopic rate constants and only under certain circumstances approximates to the dissociation constant (i.e. K_D), it is often used to measure the affinity of an enzyme for a given substrate, i.e the strength of the enzyme-substrate complex (high K_M indicates weak binding and low K_M indicates strong binding). Table 3 shows K_M and pK_M values.

Table 3. K_M [μM] values from Seeman et al. (Seemann et al., 2002). pK_M (-log K_M) are also reported.

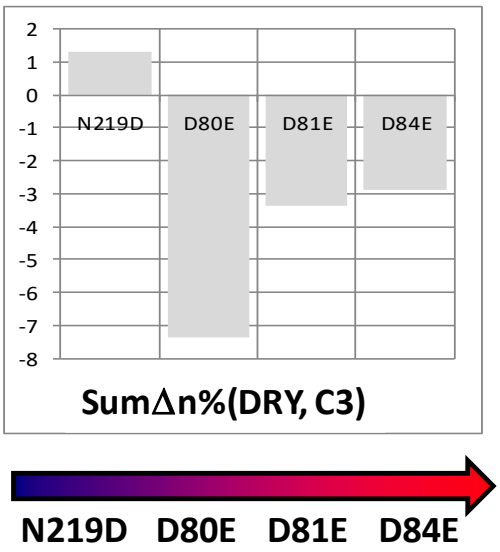
Mutant	**K_M(μM)**	**pK_M**
W308F	0.16	0.80
H309A	0.19	0.72
H309S	0.25	0.60
F77Y	0.30	0.52
H309C	0.33	0.48
H309F	1.43	-0.16
D84E	2.28	-0.36
D81E	2.60	-0.41
D80E	6.98	-0.84
N219D	16.60	-1.22

D80, D81, D84, and N219 residues have been shown to be directly involved in binding of the required divalent Mg^{2+}cofactors that interact with the pyrophosphate moiety of the substrate and are either essential or very important for both turnover activity, as experimentally measured (Seemann et al., 2002). By contrast, F77, W308 and H309 line the proposed binding pocket and are involved in less well understood hydrophobic interactions with the farnesyl portion of the substrate as well as proposed quadrupole cation interactions with the various cationic reaction intermediates (Seemann et al., 2002). This complex picture is expected to be mirrored in the computational analysis, in which a coherent trend of K_M and k_{cat}/K_M vs the computational parameters should only be verified for the four mutants directly involved in the substrate binding (see above).

Figure 3 shows the score based on C3 and DRY probes (SumΔn%(DRY,C3)) in relation with mutants activity expressed as pK_M for the four mutants directly involved in the binding of the substrate (see above).

There is a clear correlation between pK_M and this score, with the exception of N219D.

Figure 3. SumΔn%(DRY,C3) was calculated according to eq. 3 and is reported in order of increasing activity. Only mutants with residues directly involved in the binding are shown.



These results highlight the major role played in this system by steric and hydrophobic forces, and the deviant behaviour of the mutant in position 219.

Our efforts were then addressed to shedding more light on mutations at position 219. The paper by Seeman et al. (Seemann et al., 2002), from which experimental K_M values are taken, also determines N219A and N219L as inactive mutants. Since the X-Ray structure of N219A is available (pdb code: 1HM7) we compared the geometry of the active site of the two crystallographic structures (1PS1 vs. 1HM7). Interestingly, the introduction of a mutation in position 219 causes a backbone rearrangement that significantly alters the active site region (data not shown). We thus replaced 1PS1 (WT) as refined structure with 1HM7 (mutant N219A). On this latter, we applied the rotamer library method, and calculated new GRID/BIOCUBE4mf scores (i.e. SunΔn%(DRY,C3)). The results (Table 4) are in line with experimental data, since N219D is predicted to be more active than N219A and N219L.

Table 4. GRID-BIOCUBE4mf results for N219 mutants with crystal structure of the mutant N219L (pdb code: 1PS1) taken as RS. Data for N1 and O probe were not reported because of the poor number of points selected by GRID/BIOCUBE4mf methods with default thresholds.

CONFORMER	DRY	Δn%	C3	Δn%
RS (N219L)	1111		109	
L219A	1169	5.22	107	-1.83
L219Y_1	893	-19.62	123	12.84
L219Y_2	992	-10.71	127	16.51
L219Y_3	888	-20.07	148	35.78
L219Y_4	876	-21.15	154	41.28
L219Y_5	757	-31.86	139	27.52
L219Y_6	911	-18.00	153	40.37
L219Y_7	757	-31.86	171	56.88
L219Y_8	1139	2.52	146	33.94
L219Y_9	1001	-9.90	128	17.43
L219Y_10	712	-35.91	152	39.45
L219Y_11	959	-13.68	135	23.85
L219Y_12	763	-31.32	126	15.60

This finding suggests that, in the absence of WT X-ray structure, information about active site rearrangements due to point mutations could also be indirectly intersected through the deviant behaviour of mutants.

In silico prediction of site-directed mutagenesis effects could be successfully used to elucidate the mechanisms of the reaction and specificity control of the aforementioned enzyme. At present, experimental massive mutagenesis studies have been performed on the basis of sequence comparisons or structural information, thereby focusing on the selected candidate residues for systematic amino acid substitution, rather than random mutagenesis. This approach is expensive and time consuming. Application of our method could significantly reduce the number of experiments. At the same time it is expected to improve the mutants screening. Summing up it is expected to significantly contribute to the increase the production levels of terpenoids by engineering the catalytic capacity of terpene synthases.

CONCLUSION

The high cost of setting up experiments to produce and characterize mutants might be mitigated by introducing an *in silico* predictive strategy, which could reduce the time needed to investigate the effect of a particular mutation in the WT enzyme, and select variants that it is worthwhile to prepare. Mutant modeling is thus a promising field in protein engineering, but homology modeling coupled with docking is still too expensive, and also too approximate, for these purposes.

The novel, general, and convenient computational strategy described here ranks the influence of mutations on biological processes, both in quali and in quantitative terms. The starting point is the reference 3D structure (usually the WT) that can be of either computed or experimental nature. Then informations about the binding site are collected from available experimental data. After checking the WT binding site for its mobility, mutants are built by the rotamer library approach, and new descriptors ($\Delta n\%$ and $Sum\Delta n\%$) derived from the GRID/BIOCUBE4mf method are used to evaluate changes induced by a single mutation on the WT, in terms of altering the interaction pattern with the ligand. $Sum\Delta n\%$ is a simple descriptor to predict pK_M, whereas the more complex analysis of $\Delta n\%$ profiles

could shed light on mutant-ligand interaction mechanisms. The method is faster and more specifically tailored to mutation studies than is docking. Moreover, it can alert the user to the presence of mutations that cause backbone rearrangements, altering active site geometry.

Caution must also be used in applying this protocol for predictive purposes when the role of the residues in the binding mechanism is not clear (the method does not dock any ligand in the binding site).

Work is in progress to extend the applications of the approach described here to include different biosystems.

REFERENCES

Atilgan AR; Durell SR; Jernigan RL; Demirel MC; Keskin O (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.*, *80*, 505-515

Braiuca P, Cruciani G, Ebert C, Gardossi L, Linda P (2004) An Innovative Application of the "Flexible" GRID/PCA Computational Method: Study of Differences in Selectivity between PGAs from Escherichia coli and a Providentia rettgeri Mutant. Biotechnol Prog 20:1025-1031

Caron G, Nurisso A, Ermondi G (2009) How to Extend the Use of Grid-Based Interaction Energy Maps from Chemistry to Biotopics. ChemMedChem 4:29-36

Chiappori F, D'Ursi P, Merelli M, Milanesi L, Rovida E (2009) *In silico* saturation mutagenesis and docking screening for the analysis of protein-ligand interaction: the Endothelial Protein C Receptor case study. MC Bioinformatics, 10(Suppl 12):S3 doi:10.1186/1471-2105-10-S12-S3

Dunbrack RL Jr (2002) Rotamer libraries in the 21st century. Curr Opin Struct Biol 12:431-440

Eyal E, Najmanovich R, Edelman M, Sobolev V (2003) Protein Side-Chain Rearrangement in Regions of Point Mutations. Proteins 50(2):272-282

Faber HR, Matthews BW (1990) A Mutant T4 Lysozyme Displays Five Different Crystal Conformations. Nature, 348:263-266

Goodford PJ (1985) A Computational Procedure for Determining Energetically Favorable Binding Sites on Biological Important Macromolecules. J Med Chem 28:849-857

Jain AN (2009) Effects of protein conformation in docking: improved pose prediction through protein pocket adaptation. J Comput Aided Mol Des 23:355–374

Lengauer T, Rarey M (1996) Computational methods for biomolecular docking. Curr Opin Struct Biol 6:402-406

Sciabola S, Stanton RV, Mills JE, Flocco MM, Baroni M, Cruciani G, Perruccio F, Mason JS (2010) High-Throughput Virtual Screening of Proteins Using GRID Molecular Interaction Fields. J Chem Inf Model 50:155–169

Seeback B, Reulecke I, Kaemper A, Rarey M (2008) Modeling of metal interaction geometries for

protein-ligand docking. Proteins 71:1237-1254

Seemann M, Zhai G, de Kraker JW, Paschall CM, Christianson DW, Cane DE (2002) Pentalenene Synthase. Analysis of Active Site Residues by Site-Directed Mutagenesis. J Am Chem Soc 124:7681-7689

Soss M. (2003) Rotamer Exploration and Prediction. http://www.chemcomp.com/journal/rotexpl.htm

Spyrakis F, Amadasi A, Fornabaio M, Abraham DJ, Mozzarelli A, Kellogg GE, Cozzini P (2007) The consequences of scoring docked ligand conformations using free energy correlations. Eur J Med Chem 42(7):921-33

Vasquez M (1996) Modeling sidechain conformation. Curr Opin Struct Biol, 6:217-221

Wade RC, Clark KJ, Goodford PJ (1993) Further Development of Hydrogen Bond Functions for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure . 1. Ligand Probe Groups with the Ability To Form Two Hydrogen Bonds. J Med Chem 36:140-147

Wade RC, Goodford PJ (1993) Further Development of Hydrogen Bond Functions for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure . 2. Ligand Probe Groups with the Ability To Form More Than Two Hydrogen Bonds. J Med Chem 36:148-156

Zhang Y (2008) Progress and challenges in protein structure prediction. Curr Opin Struct Biol 18:342-348