Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 25 (2007) 671–679
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Empirical rules facilitate the search for binding sites on protein surfaces

Henrik te Heesen [a], Anna Melissa Schlitter [b], Jürgen Schlitter [a,*]

[a] *Biophysics Department, ND 04 Nord, University of Bochum, 44780 Bochum, Germany*
[b] *School of Medicine, University of Bochum, 44780 Bochum, Germany*

Received 15 November 2005; received in revised form 11 April 2006; accepted 8 May 2006
Available online 12 May 2006

Abstract

Computational surface screening of 3D protein structures is a valuable means of finding possible docking sites for substrates, effectors and similar molecules. It can be improved by considering properties of molecules which are known to bind to protein surfaces, and thus reflect the required properties of binding sites. In-depth studies are available on drugs and lead compounds as binding partners with statistically assured properties. Here we present a simple strategy for finding binding sites, which is based on the empirical rule-of-five by Lipinski et al. for oral drugs and the rule-of-three by Congreve et al. for leads. The fast automated search with the new C-code TRIDOCK yields a preliminary set of sites, thus facilitating further investigation by visual, comparative and quantitative work. Possible binding sites are tagged by pseudo-atoms added to the structure file for molecular graphical evaluation. Usually, the strategy yields not just a few single sites, but an accumulation of several sites in known substrate binding pockets. Clusters are also found at known or putative protein–protein docking interfaces. A comparison of the activated and inactivated form of the GTPase Ras reveals clear differences and identifies a niche, which is possibly a suitable new target for compounds that bind specifically to activated Ras.

Keywords: Proteins; Enzymes; Binding sites; Docking; Lipinski's rule-of-five; RO5; Lysozyme; Ras; Ras–Raf interaction; Inhibition; Drugs; Lead compounds

1. Introduction

Proteins are knots of a complicated network of interactions which, apart from the interaction with small diffusing molecules, are mediated by inter-protein surface contacts of varying stability. The detection of docking sites and docking modes for known or hypothetical compounds is the focus of numerous computational procedures [1–3]. There are also codes for finding internal cavities, channels or surface invaginations without specifying a corresponding binding partner, e.g. SPHGEN [4], Surfnet [5], PASS [6], or CASTp [7]. These codes give a complete list of geometrical features, partially including volume and area. Of course, in the next step binding properties of the protein surface have to be taken into account [6]. This aspect is the focus of the current study. It aims to support the visual inspection of a protein surface – always the first approach to a new structural model – by monitoring surface features which could be envisaged as binding sites for small compounds. It is expected – and will be demonstrated – that such patterns can belong to the much larger interface used by docking proteins.

We shall start by considering small compounds because a statistically confirmed set of properties is available for them. The rule-of-five (RO5) by Lipinski et al. [8,9] for oral drugs and the rule-of-three (RO3) by Congreve et al. [10] for leads are simple enough to provide a fuzzy, but all-embracing definition of relevant compounds. Molecules are described as lead-like when they bind with relatively weak affinities in the high micromolar to millimolar range. Here the definition of binding compounds is transformed into an also fuzzy definition of a complementary binding site. The resulting rules for the selection of binding sites on a protein cannot be free of parameters which have to be tailored for generating samples that are large enough to cover known sites, yet small enough to handle. A strategy is developed to deal with these rules, and converted into a fully automated fast C-code named TRIDOCK. It was validated on several enzymes which were co-crystallised with drugs, but only after removing the drugs.

* Corresponding author.
E-mail address: juergen.schlitter@rub.de (J. Schlitter).

1093-3263/$ – see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2006.05.005

The examples shown and discussed below are further proteins with known substrates, ligands and protein binding partners. All structures are taken from the Protein Data Bank (pdb) [11].

Special attention is given to the GTPase Ras which interacts with several proteins [12] and is considered a possible target for drugs [13]. We use this example to probe and discuss how, after automated pre-selection, visualisation and comparison with related structures can confirm the prominent role of a conspicuous niche. Moreover, we discuss the structure and role of possible lead compounds which can be anchored in this niche.

2. Methods

The RO5 summarised in Table 1 provides the upper limits for some parameters describing binding compounds. Their distributions exhibit maximum probability near about one half of the limits. The RO3 for the more weakly binding leads roughly indicates limits below which binding affinity rapidly drops. Among the properties considered in the rules, the number of rotatable bonds (e) and lipophilicity (b) provide little information about a binding site. We therefore evaluate only the molecular weight and the hydrogen bonding capacity, which are transformed into requirements for binding sites.

Properties (c) and (d) suggest the presence of at least three H-bonds. Therefore only triangles spanned by nitrogens (N) and/or oxygens (O) on the protein surface are the features that will be scanned (our rule 1, R1).

The molecular weight (a) provides an insight of areas and typical lengths. Starting from the mass density (mass per area) of benzene, $\rho = 5.1$ D/Å^2, the mass limits of 300–500 D correspond to 60–100 Å^2 (our R2). For the sake of efficiency, it seems reasonable to make a pre-selection on the level of distances between donor/acceptor atoms. Equilateral triangles meeting the area criterion have side lengths of 12–15 Å. Distances of less than 4 Å can be excluded due to the H-bond geometry. Extreme distances can be excluded because large compounds tend to have more than three donors/acceptors with a lesser distance. The resulting interval is thus 4–12 Å (our R3).

The next rule claims that a compound, and hence also the representative triangle, is embedded in a niche or pocket (our R4). This is clear for substrates or receptor ligands and compounds which act as inhibitors to replace such molecules. Pockets are also known to contribute to protein–protein docking where indenting of protrusions and invaginations is needed for recognition and stability.

To satisfy the 'pocket criterion' R4, we first discard all pairs of donors/acceptors which are connected by a line that runs through the protein. Halfway between the pairs of donors/acceptors, we therefore place a small probe sphere as shown in Fig. 1. We reject a pair if the sphere collides with a protein atom. This is our rule R4a. Then triangles are constructed from the remaining pairs and a large probe sphere is placed at the centre of each area. Again, we discard the triangle if the large sphere collides with a protein atom, which is our rule R4b. The large spheres are stored as tags for possible binding sites.

Table 1
Properties claimed by Lipinski's and Congreve's rules

	RO5 (Lipinski: oral drugs)	RO3 (Congreve: leads)
(a) Molecular weight	≤500	≤300
(b) Lipophilicity log P	≤5	≤3
(c) Hydrogen donors	≤5	≤3
(d) Hydrogen acceptors	≤10	≤3
(e) Rotatable bonds	–	≤3

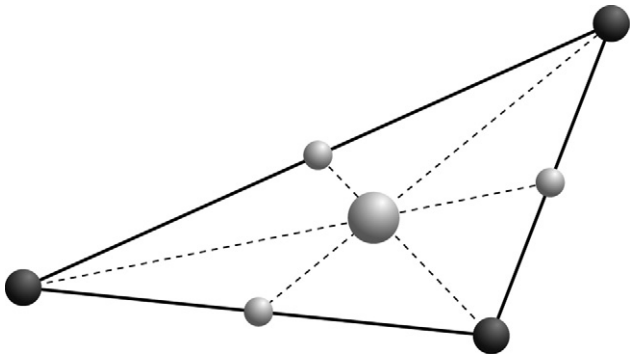


Fig. 1. Triangle spanned by three oxygen or nitrogen atoms (black) marking possible H-bond donor or acceptor sites. Probe spheres are placed on side and area centres to check the distance from the protein surface (pocket criterion). When a triangle satisfies our set of rules, it is accepted as a possible binding site and the central sphere is retained as a tag.

The numerical algorithm to find possible docking positions consists of three steps:

(i) Find donor and acceptor atoms on the protein surface (R1), cf. Fig. 2a.
(ii) Build a distance list containing all pairs of donors/acceptors satisfying the distance criterion (R3) and the pocket criterion for triangle sides (R4a), cf. Fig. 2b.
(iii) Use accepted pairs to create triangles satisfying the pocket criterion for triangle centres (R4b) and the area criterion (R2), cf. Fig. 2c and d.

The numerical realisation requires introduction of a few parameters. The admissible intervals for triangle areas, 60–100 Å^2, and for side lengths, 4–12 Å^2, were determined above. In order to find the donor and acceptor atoms on the protein surface, the protein is embedded in a 3D grid. The grid points have a distance of $d_{\text{grid}} = 1.0$ Å. All grid points within 3.0 Å of the protein are deleted. The donors and acceptors on the surface are found by selecting all oxygen and nitrogen atoms as docking candidates within 3.5 Å of the remaining grid points (R1). For the pocket criterion, the small probe spheres on the side centres are given a radius of 1.9 Å (R4a); the large spheres on the area centres have a radius of 2.5 Å (R4b). The radii are minimum distances from the centres of protein atoms.

The procedure is applied to pdb files after removing all hydrogen atoms and solvent molecules. Other hetero molecules can optionally be retained. The standard output is the original pdb file completed by the tags stored as pseudo-atoms. They can also be used for highlighting binding sites. In addition, if desired, the corresponding donor/acceptor atoms and triangles can also be obtained.

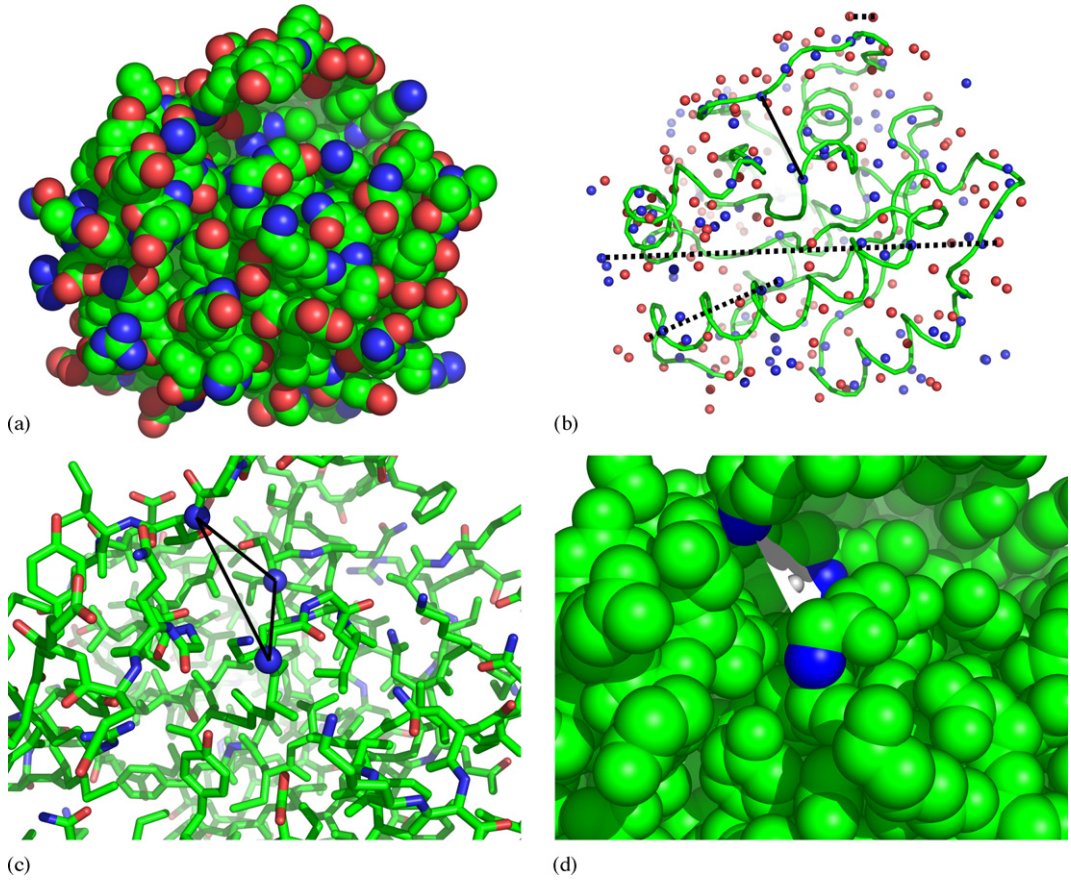


Fig. 2. Essential steps of the search for binding sites. (a) Oxygen (red) and nitrogen atoms (blue) are determined on the protein surface. (b) Distance lines are analysed and accepted (black line) or discarded (dashed lines) if they are too short, too long, or run through the protein. (c) Triangles are constructed. (d) The resulting triangles are retained only if their area is admissible and a pocket is located there. The central sphere can be used to highlight the site as shown below.

The code was developed as a PyMOL script, which allows immediate inspection of results, and is now available as a C-code named TRIDOCK. Screening a 20 kDa protein takes 10 s on a 1.3 GHz PC. The computing time is very sensitive to the initial distance criteria. A decrease of the grid distance d_{grid} from 1 to 0.5 Å or increase of the admissible side length l_{max} from 12 to 15 Å causes a factor of 5–10 in computing time. The number of triangles, however, merely increases by about 10%. More importantly, no additional clusters are found. Clusters, not single triangles will be shown to be typical of a functional binding site. The additional tags tend to fill up the clusters which are already present. Obviously, the algorithm is robust with respect to the parameters mentioned.

3. Results

All test runs with enzyme structures which had been obtained from co-crystallised protein–drug complexes showed that the code reliably finds the binding pocket. Tests were carried out on phospholipase A2 co-crystallised with the small 181 Da acetylsalicylic acid (1OXR) and other cases up to phosphodiesterase co-crystallised with the large 666.7 Da sildenafil citrate (1XOS). The results show that the rules for binding sites satisfactorily reflect the features of drugs. More precisely, tags are found only in the hydrophilic part of the binding pockets, which in the examples is clearly separated from the adjacent hydrophobic part which, according to the lipophilicity stated in the RO5, must always be existent there. Hydrophilic pockets are also tagged when they are large enough to contain several water molecules. In all cases, they are highlighted not by a single tag, but by a cluster of tags. Another general feature of proteins seems to be their possession of several hydrophilic pockets or niches on the surface. The simplest interpretation would be their contribution to solubility. The next examples of proteins for which there are no known complexes with drugs, but instead with other interaction partners, will help to better understand the findings just mentioned.

3.1. Example of lysozyme

Lysozyme is a small 129 residue-long cytoplasmic glycosidase cleaving polysaccharides like murein found in cell walls and chitin [14]. The substrates bind in an almost linear conformation in a cleft on the protein's surface. It exhibits a clear separation of a binding pocket for four monomers and the adjacent proper catalytic site, where the ionised Asp52 and the un-ionised (protonated) Glu35 are the essential conserved residues. As a consequence, short chains are bound, but not cleaved and can act as inhibitors. The cleft is

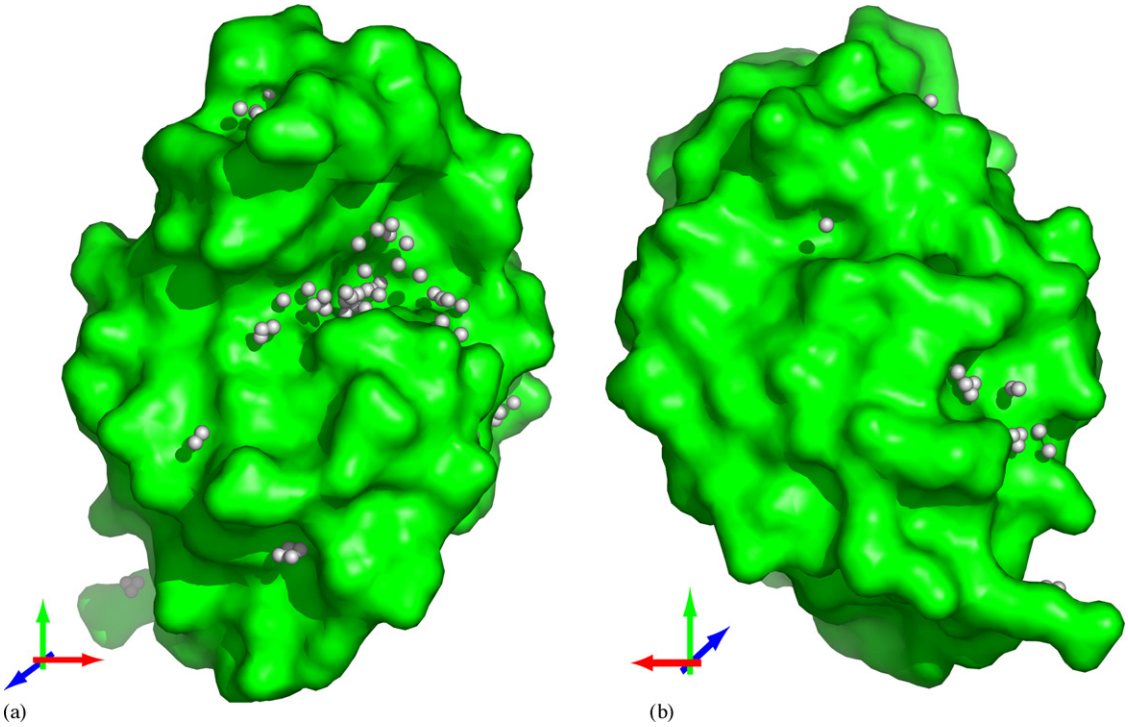


Fig. 3. Lysozyme with substrate binding pocket ((a) front view and (b) back view). The white tags are centres of triangles of H-bond donors/acceptors filtered out by the set of rules explained in Section 2. The underlying pdb file is 3LZT.

seen in the upper half of Fig. 3a where it runs from the binding pocket (right-hand side) through a bottleneck (middle) where the carboxylates are situated before it opens to the left. Note that the underlying structure 3LZT was obtained from lysozyme alone, i.e. not co-crystallised with a substrate analogue.

The screening procedure positions a clear track of tags along the known cleft, each tag representing a triangle of H-bond donors or acceptors. The accumulation of tags is a predictable result: the small trimeric inhibitor tri-NAG already forms 6 H-bonds to the protein [14], which results in 20 triangles. A smaller cluster is found at the right-hand side. Adjacent residues belong to the interface of a monoclonal antibody docking below, as can be seen from the crystal structure of the complex, 1G7M. By reorientation of side chains, however, the niche changes completely. Most of the reverse is sparsely populated by tags. Only on the right is there a flat niche which highlighted by a cluster. Its functional role is unknown.

Lysozyme nicely reflects the added information obtained by the method. When displayed with a simple protein viewer, its surface shows an unstructured distribution of a total of 225 nitrogens and oxygens, i.e. possible acceptors and donors. The number of triangles or tags filtered out of the possible triplets ($>10^6$) by application of the criteria is less, in this case only 71. The essential gain is their concentration to a few clusters, which are the most plausible candidates for real docking sites.

Most clusters exhibit adjacent aromatic and other hydrophobic residues, which allow docking of partner molecules with hydrophobic components. This seems to be a general feature of protein surfaces which should not be regarded as an extra criterion for surface patches, but is relevant for the binding molecules. Likewise, low internal flexibility of a compound probably implies not a condition for the protein, but for the compound itself which binds better with a smaller loss of entropy. This consideration justifies the neglect of the properties (b) and (e) of RO3 and RO5 shown in Table 1.

On the other hand, we had added the condition that a slight impression on the protein surface should be there. The binding sites of lysozyme are indeed only found in concave regions, but not all niches or pockets meet the conditions. Some seem to be too small, others like the one at the bottom left on the front of lysozyme, are too hydrophobic.

3.2. Example of bacteriorhodopsin

Bacteriorhodopsin is a seven helices transmembrane protein functioning as a light-driven proton pump in *Halobacterium salinarum*. The analysis was performed on the 248-residue long X-ray structure 1QHJ [15]. As shown in Fig. 4, no binding sites are determined on the surface of the transmembrane region, which is roughly identical to the helical portion. Although the surface there is not smooth, but uneven with many niches, the sparse distribution of donors/acceptors does not seem suitable for the anchoring of hydrogen-bonding compounds.

The two clusters on the solvent-accessible sides (above and below) do not belong to known docking sites. Instead, they mark the hydrophilic funnels of the entrance and exit channels, as recently shown by molecular dynamics simulations of water movements [16,17]. It is an unexpected feature that the funnels possess characteristics of binding sites for compounds that are considerably larger than hydronium H_3O^+. One may speculate that the funnels are adapted to larger structures such as the 'Zundel cation' $H_5O_2^+$ or the 'Eigen cation' $H_9O_4^+$, which are

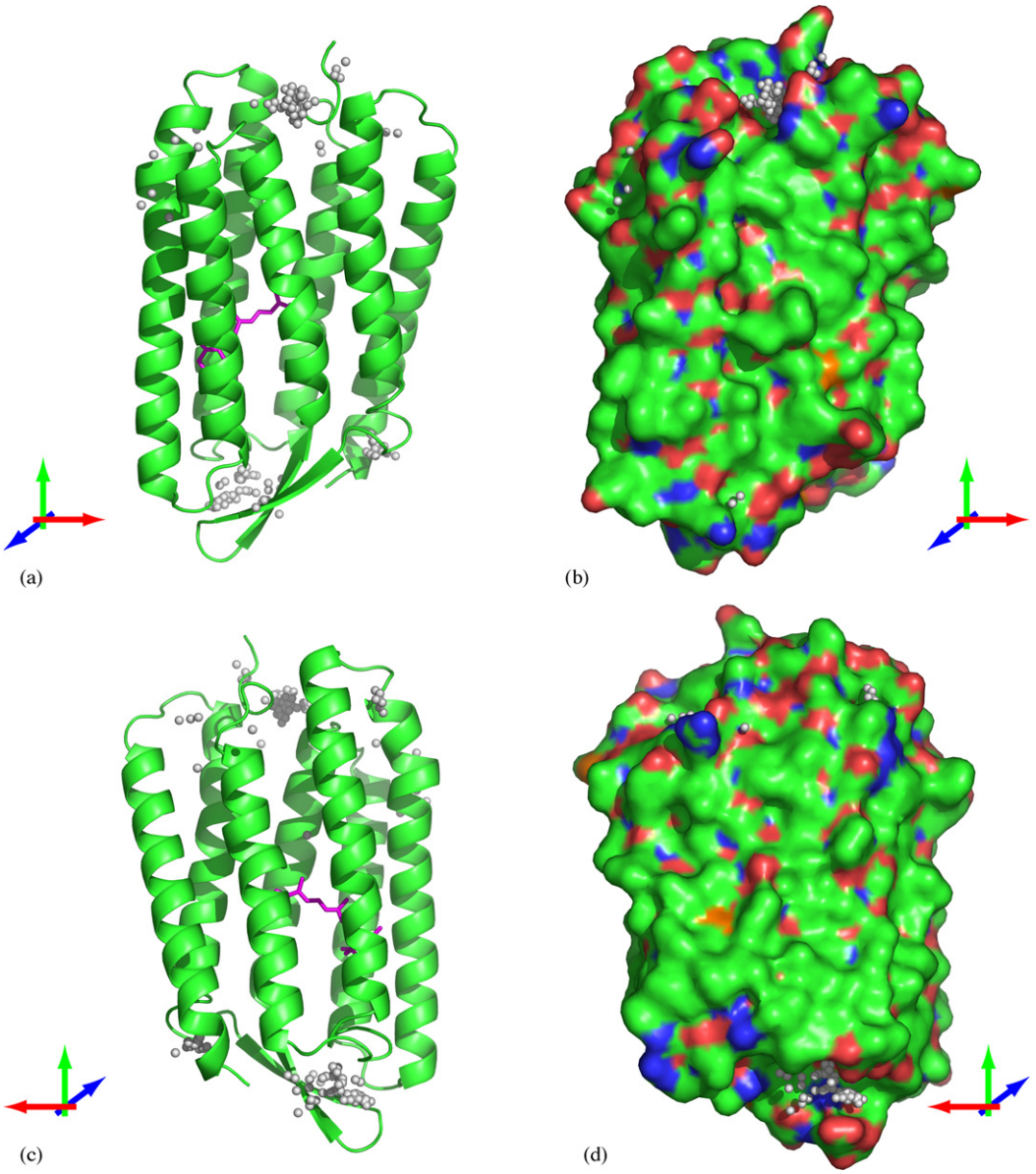


Fig. 4. Bacteriorhodopsin ((a and b) front view and (c and d) back view). The protons enter from the cytoplasmic side (above) and are pumped to the periplasmatic side (below). The underlying pdb file is 1QHJ. On the Connolly surface, nitrogens are shown in blue and oxygens in red. Retinal is shown in magenta.

discussed in the literature [18] as elements of proton conduction in water. The idea of protons being transported through hydrogen-bonded water networks inside bacteriorhodopsin has been discussed previously, and was recently proven by means of experiments [19].

The example bacteriorhodopsin also shows a limitation of our method when applied with the given parameters. The above methods for finding pockets and cavities actually find the small internal cavities inside bacteriorhodopsin which contain only a few water molecules and contribute to proton transport. The parameters of the current method were chosen with the aim of finding larger pockets or cavities suited to drugs or larger groups. Our tests show that the cavities for functional water in bacteriorhodopsin require a finer solvent grid with $d_{\text{grid}} = 0.5$ Å and a milder pocket criterion with $r_{\text{area}} = 1.9$ Å.

3.3. Example of Ras

Ras is an example of guanosine nucleotide-binding proteins which – apart from the substrate guanosine triphosphate (GTP) – are known to interact with many other proteins [12] and small compounds [20,21]. H-Ras, the variant shown below, has 189 residues. A terminal tail of about 20 residues is not contained in the structures. Ras is a target for drugs because of its role as a switch in the signal transduction pathway controlling cell proliferation [13]. It is activated (switched on) by catalysed uptake of GTP and deactivated (switched off) by hydrolysis of GTP to the diphosphate GDP, which is mostly catalysed by a further protein (GAP). Only the activated form allows docking of a downstream effector, which is the next step in signal transduction. Thus signal transduction via Ras in healthy cells is regulated and restricted to a short-term interval. Functioning

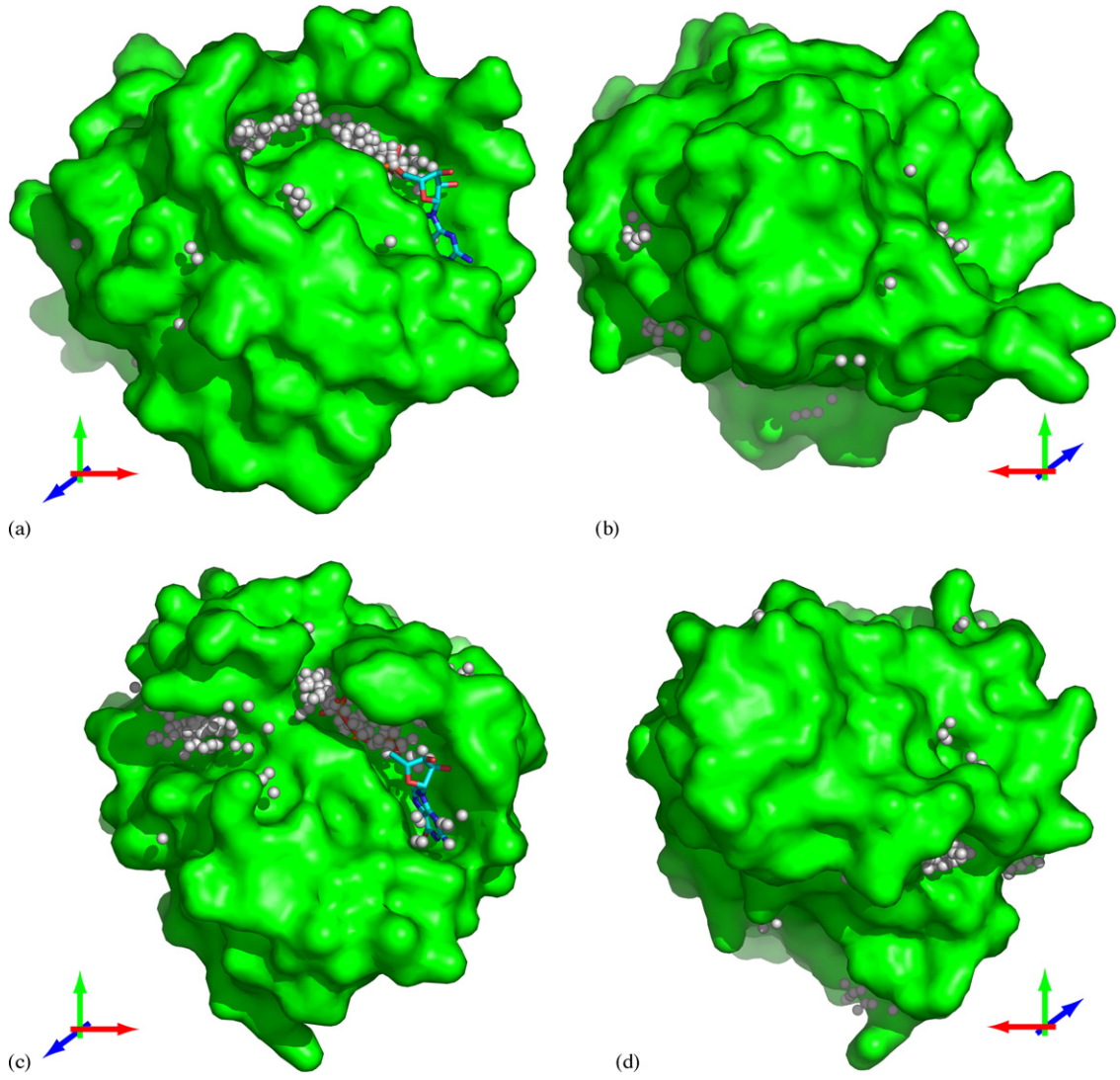


Fig. 5. Front and back view of the inactive Ras-GDP conformation (top, pdb file 1Q21) and activated Ras-GTP conformation (bottom, 1QRA). The substrates were not regarded during screening, but inserted later. On the front (a and c), the tags show the binding pockets and a new binding site (c, middle left), which emerges after activation. In both conformations the reverse of the binding pocket is rather flat. Indications of binding sites are found at the middle left and bottom left.

hydrolysis is indispensable for regulation. Tumorigenic mutants of Ras remain in the activated Ras-GTP conformation, which interacts with the Raf kinase as the next partner in the signal chain, thus resulting in an uncontrolled permanent signal.

The inactivated Ras-GDP form shown in Fig. 5a has a large binding pocket containing GDP. Our method finds three clusters of binding sites in a long cleft one of which covers the diphosphate binding pocket, and none at the guanosine which is indeed essentially stabilised by hydrophobic contacts. The central cluster highlights the binding site for the gamma phosphate, which is cleaved during deactivation and therefore not occupied here. The cluster to the far left indicates the existence of an extra binding niche with an unknown function. There are no indications for other binding sites on the front of Ras-GTP. The reverse exhibits only one conspicuous cluster of tags in a wide inversion at the bottom left. In the related Ran-Importin Beta complex (1IBR), this region is part of the common interface.

The front view of the activated form Ras-GTP in Fig. 5c reveals a very different conformation. The extra niche has disappeared and a new cleft with a cluster of tags has opened at the top left. This becomes clearer in Fig. 6 where adjacent loops are given different colours. The new pocket lies under the switch II loop (magenta) and is part of the docking interface of the activator protein GAP [22] and the exchange factor SOS [23]. GTP is embedded in the P-loop (cyan) and covered by the switch I loop (yellow). Behind GTP there is a channel-like rear exit to the surface, which will now be discussed in more detail.

Note that the results shown so far were obtained from the enzymes with the ligands removed. By way of comparison, we also performed a surface screening of Ras-GTP with GTP in the pocket. Fig. 7 shows a view of the binding pocket, which was optimised to show the residual binding capacity when GTP is present. A cleft above the gamma-phosphate is highlighted by ten tags and is accessible to a flat compound. For instance, experiments demonstrated that Zn^{2+}-cyclen, a small metal

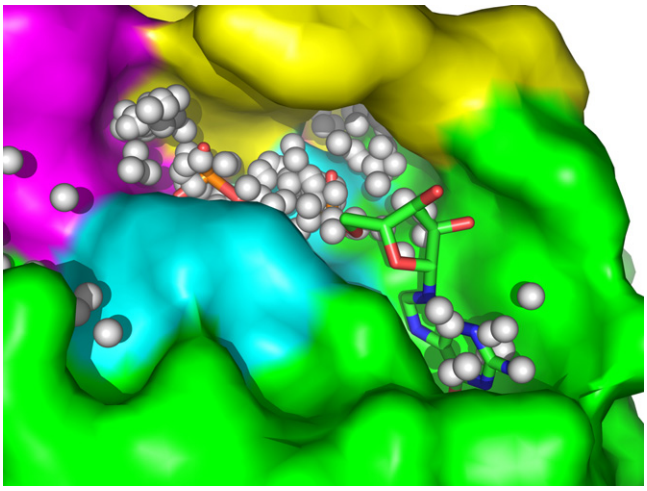


Fig. 6. GTP binding pocket and neighbouring area in activated Ras. GTP was inserted after screening. The switch I loop (yellow) and switch II loop (magenta) take their activated conformation with a new pocket on the left and a distal niche opening behind GTP under switch I. The predicted binding sites exhibit a distribution which differs from the inactivated case.

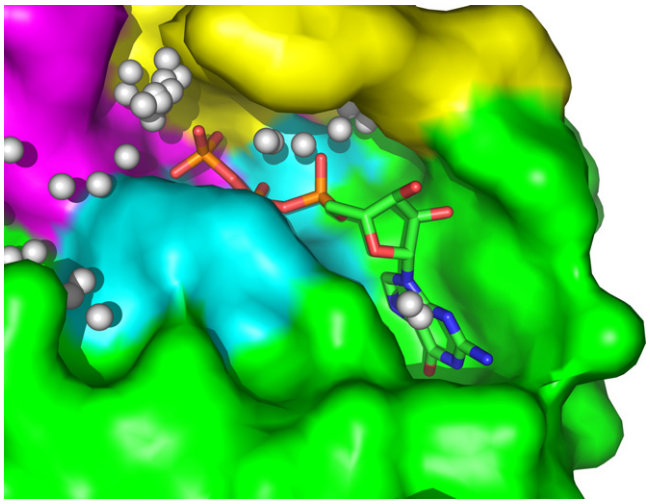


Fig. 7. Binding pocket of activated Ras with substrate GTP screened with substrate. The hydrogen bond capacities indicated in Fig. 6 are now utilised by GTP. A residual binding site containing ten tags becomes visible in a cleft above the gamma-phosphate. The view is similar to Fig. 6, but optimised for the cleft.

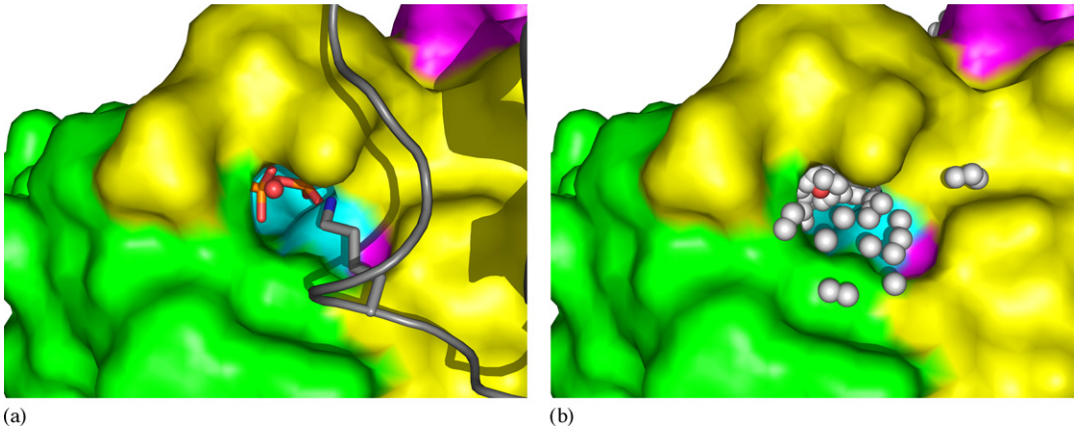
chelate, binds to the activated conformation of Ras. The exact position could not yet be identified, but interaction with the triphosphate was detected [21]. In addition, it was observed that a sodium ion moved to a stable position at the gamma-phosphate in a molecular dynamics simulation [24]. In summary, the activated Ras structure 1QRA exposes 288 acceptors/donors to the solvent when the substrate is removed. They result in 277 tags which, however, accumulate in much less functional clusters.

3.4. Distal binding niche

All kinds of activated Ras-GTP, whether alone or in a complex with activator or effector proteins or after replacing GTP by a non-cleavable analogue, exhibit a conformation of the effector loop which – in contrast to inactivated Ras-GDP – opens a niche from the surface down to the alpha-phosphate. It is not viewed from behind in Fig. 5, but from above as shown in Fig. 8. It is identical to the rear exit of the GTP binding niche observed in Fig. 6. The screening procedure reveals its binding capacity by placing a cluster of almost the same number of tags as found for the triphosphate pocket. Since the niche is partially formed by the reverse of the effector loop, which is not in contact with the GTP substrate, it is termed the distal binding niche (DBN).

The conformational change of the effector loop during deactivation was recognised at an early stage and gives an explanation for signal transduction by docking of the Raf kinase, which occurs only once Ras is activated. The effector loop is part of the interface in the structural model of the Ras–Raf binding domain complex [25], where a lysine forms a H-bond to Asp33 on the surface without intruding in the niche.

However, there is only one X-ray structure (1WQ1 [22]) where the binding capacity of the DBN is utilised. The Lys949 side chain of the activator docks to Ras and makes an ionic contact to Asp33 and intrudes deeply into the pocket. It displaces most of the water seen previously. The 5.65 Å gap



(a) (b)

Fig. 8. (a) View of the distal binding niche (DBN) in the Ras-GAP complex (pdb file 1WQ1). Lys949 of GAP protrudes into the niche and is connected with the alpha-phosphate via a water molecule, W217. The residue is part of a loop (grey), which runs over the DBN. (b) Screening the surface after removing GAP reveals the conspicuous binding capacity of the DBN.

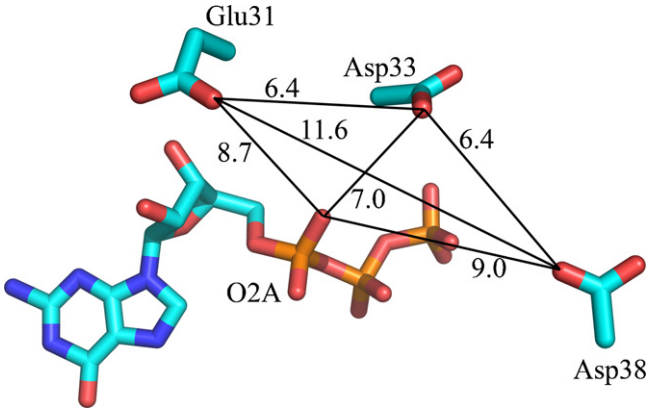


Fig. 9. Schematic presentation of the distal binding niche as a hollow tetrahedron spanned by four charged groups. The distances (in Å) vary by 0.7 among known structures. The niche is open only in the activated Ras-GTP conformations.

between Nε of Lys949 and the non-bridging oxygen O2A of the substrate is bridged by a water molecule. This binding mode shown in Fig. 8 was not found for effectors, but the activator GAP-334 [22].

Schematically the DBN is a hollow tetrahedron framed by four charged H-bond acceptors depicted in Fig. 9. Three carboxylic side chains – Glu31, Asp33 and Asp38 – mark the outer edge, while an almost buried, but water-accessible non-bridging oxygen (O2A) of GTP is found on the innermost position. It is partially hydrophilic and contains at least six water molecules [26], connecting GTP with the bulk water as long as no other molecule docks at the site. The distances of the characteristic acidic residues show little variation among the activated conformations available in the Protein Data Bank, but change drastically during deactivation. Moreover, the conformational change during deactivation rotates Tyr32 so that the niche is closed.

4. Discussion

The new screening procedure TRIDOCK was applied to very different proteins in order to test its capacity to make a preselection of possible binding sites. In enzymes co-crystallised with drugs, the hydrophilic parts of the binding pockets are always found. This is not surprising as the criteria for pockets are derived from drugs and leads. For lysozyme (crystallised without substrate) the known substrate niche was found while most of the surface was discarded and the same applied to several Ras structures (crystallised with substrate or ligands). Indications for additional binding sites were found in both cases, and their function could partially be elucidated by comparison with crystal structures of relevant complexes. This supports the idea that drugs and leads make use of the same binding sites which are used in functional contacts, e.g. by other proteins. The results thus prove our approach to be a valuable method for the envisaged purpose, namely supporting the always needed visual inspection in the search for functional parts of the protein surface.

Among the sites discovered on activated Ras, particular attention should be paid to the distal binding niche. (a) The activator GAP inserts its Lys949, a conserved residue, so deeply in the DBN that the (positively charged) amino group touches the water molecule W217 which is in contact with the (negatively charged) oxygen of the alpha phosphate of GTP, cf. Fig. 8. This suggests a certain role in catalysis. In fact, the mutation Lys to Ala at the corresponding position in the activator NF1 [27] reduces k_{cat} by 30%. The increase of K_D was even greater (factor 10). (b) Any compound that binds in the DBN and has at least the size of a lysine residue obviously represents a competitive inhibitor of Ras–Raf interaction. The appendix contains proposals for compounds which are conceivable leads for that purpose. In constitutionally activated oncogenic Ras, it can interrupt the uncontrolled signal for cell division [13].

Potential binding sites for small leads can be determined by this algorithm, but this question was not the focus of this work. Clusters of sites were shown to be real binding sites of substrates, inhibitors or other proteins. Of course, only pocket-like patches are found which contribute to inter-protein contacts. The complementary protrusions or hydrophobic patterns are not highlighted. The clustering effect of tags was explained by the example of the tri-NAG inhibitor as a consequence of multiple H-bond triplets. Apart from clusters, a sparse distribution of non-clustered sites is mostly tagged. These are probably all ‘positive false’, i.e. do not correspond to real binding sites.

Other existing methods [4–7] mostly generate a more complete list of pockets and cavities. The underlying algorithms, however, are purely geometrical. By way of comparison, we have applied PASS [6] to bacteriorhodopsin (1QHJ) and Ras-GTP (1QRA). PASS yields a similar graphical output which simplifies the visual comparison. For bacteriorhodopsin, it identifies the same entrance and exit funnels as our tool TRIDOCK, but an additional six pockets in the transmembrane interface. Here they are automatically discarded due to poor H-bonding capacity. By PASS one additional hydrophobic pocket is found on Ras-GTP, but the distal binding niche, which was extensively discussed above, is not found. This is an example of where our code is more sensitive. Obviously, the advantage of our more functional approach is its ability to detect possible binding sites on protein surfaces – without any other specification – for compounds with three or more H-bonds and to reject other niches or pockets. Our parameters were chosen so that pockets for very small compounds, such as one to three waters, are not monitored. However, they can be adapted to also include such small objects. The increase in information is due to the selection of a few hot spots which are worth further analysis. Possibly a combination of TRIDOCK with other tools would be useful for detecting pharmacophores because they have to provide pockets with both hydrophilic and hydrophobic parts.

Appendix A

A.1. Peptidic lead binding to the DBN

A first idea of a peptidic lead is derived from the activator protein RasGAP p120. Lys949 shown in Fig. 8a belongs to a

loop with sequence "Lys Ser Val Gln Asn Leu Ala Asn Leu Val Glu Phe Gly Ala Lys Glu" covering the residues 935–950 of GAP. The first lysine and the last glutamine close the loop by an ionic H-bond of their side chains. This loop or sections of it are conceivable peptidic leads that can be anchored in the DBN. The following modifications are suggested to improve stability and affinity:

(a) Extension of the Lys949 to lysine-epsilon-*N*-methylimine or lysine-epsilon-*N*-ethylimine.
(b) Cyclisation e.g. via the side chains or a peptide bond between the first and last residue.
(c) Conservative substitution of residues.

The essential step will be the modification of Lys949 to make it a side chain which can immediately bind to GTP (without intermediate water W217) by an ionic H-bond.

A.2. Non-peptidic leads

Non-peptidic compounds should possess amino or imino groups as hydrogen donors. Examples are Putrescin, Spermidin and Spermin. Related compounds have the structure H_2N–$(CH_2)_n$–NH_2 and H_2N–$(CH_2)_n$–NH–(C=NH)–NH_2 with n between 0 and 7. A further generalised structure consists of four functional segments, R3–D2–LK–D1, where D1 is the first hydrogen donor designed to bind via the distal niche to the GTP oxygen O2A, and LK a linker ensuring the correct distance of the following group D2. This group contains one or more hydrogen donors capable of forming H-bonds to one or more of the three carboxylic side chains of Glu31, Asp33 and Asp38. Finally, R3 is a group that is capable of providing further polar and nonpolar contacts to the protein surface. In view of property (b) of Lipinski's RO5, it has to reduce solubility in order to compensate the hydrophilic groups D1 and D2 by offering hydrophobic surface patterns.

References

[1] S. Jones, J.M. Thornton, Analysis of protein–protein interaction sites using surface patches, J. Mol. Biol. 272 (1997) 121–132.
[2] S. Jones, J.M. Thornton, Prediction of protein–protein interaction sites using patch analysis, J. Mol. Biol. 272 (1997) 133–143.
[3] I. Halperin, B.Y. Ma, H. Wolfson, R. Nussinov, Principles of docking: an overview of search algorithms and a guide to scoring functions, Proteins-Structure Funct. Genet. 47 (2002) 409–443.
[4] I.D. Kuntz, J.M. Blaney, S.J. Oatley, R. Langridge, T.E. Ferrin, A geometric approach to macromolecule–ligand interactions, J. Mol. Biol. 161 (1982) 269–288.
[5] R.A. Laskowski, Surfnet—a program for visualizing molecular-surfaces, cavities, and intermolecular interactions, J. Mol. Graph. 13 (1995) 323–330.
[6] G.P. Brady, P.F.W. Stouten, Fast prediction and visualization of protein binding pockets with PASS, J. Comput-Aided Mol. Des. 14 (2000) 383–401.
[7] T.A. Binkowski, S. Naghibzadeg, J. Liang, CASTp: computed atlas of surface topography of proteins, Nucl. Acid Res. 31 (2003) 3352–3355.
[8] C.A. Lipinski, Lead- and drug-like compounds: the rule-of-five revolution, Drug Discov. Today: Technol. 1 (2004) 337–341.
[9] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Deliv. Rev. 23 (1997) 3–25.
[10] M. Congreve, R. Carr, C. Murray, H. Jhoti, A rule of three for fragment-based lead discovery? Drug Discov. Today 8 (2003) 876–877.
[11] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucl. Acids Res. 28 (2000) 235–242.
[12] I.R. Vetter, A. Wittinghofer, Signal transduction—the guanine nucleotide-binding switch in three dimensions, Science 294 (2001) 1299–1304.
[13] A. Wittinghofer, H. Waldmann, Ras—a molecular switch involved in tumor formation, Angew. Chem. Int. Ed. 39 (2000) 4192–4214.
[14] L. Stryer, Biochemistry, Freeman, San Francisco, 1995.
[15] H. Belrhali, P. Nollert, A. Royant, C. Menzel, J.P. Rosenbusch, E.M. Landau, E. Pebay-Peyroula, Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 Å resolution, Structure 7 (1999) 909–917.
[16] C. Kandt, J. Schlitter, K. Gerwert, Dynamics of water molecules in the bacteriorhodopsin trimer in explicit lipid/water environment, Biophys. J. 86 (2004) 705–717.
[17] C. Kandt, K. Gerwert, J. Schlitter, Water dynamics simulation as a tool for probing proton transfer pathways in a heptahelical membrane protein, Proteins-Structure Funct. Bioinform. 58 (2005) 528–537.
[18] N. Agmon, The Grotthus mechanism, Chem. Phys. Lett. 244 (1995) 456–462.
[19] F. Garczarek, K. Gerwert, Functional waters in intraprotein proton transfer monitored by FTIR difference spectroscopy, Nature 439 (2006) 109–112.
[20] O. Muller, E. Gourzoulidou, M. Carpintero, I.M. Karaguni, A. Langerak, C. Herrmann, T. Moroy, L. Klein-Hitpass, H. Waldmann, Identification of potent Ras signaling inhibitors by pathway-selective phenotype-based screening, Angew. Chem. Int. Ed. 43 (2004) 450–454.
[21] M. Spoerner, T. Graf, B. Konig, H.R. Kalbitzer, A novel mechanism for the modulation of the Ras–effector interaction by small molecules, Biochem. Biophys. Res. Commun. 334 (2005) 709–713.
[22] K. Scheffzek, M.R. Ahmadian, W. Kabsch, L. Wiesmuller, A. Lautwein, F. Schmitz, A. Wittinghofer, The Ras–RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants, Science 277 (1997) 333–338.
[23] P.A. Boriack-Sjodin, S.M. Margarit, D. Bar-Sagi, J. Kuriyan, The structural basis of the activation of Ras by Sos, Nature 394 (1998) 337–343.
[24] M. Klähn, J. Schlitter, K. Gerwert, Theoretical IR spectroscopy based on QM/MM calculations provides changes in charge distribution, bond lengths and bond angles of the GTP ligand induced by the Ras–protein, Biophys. J. 88 (2005) 3829–3844.
[25] M. Nassar, G. Horn, C. Herrmann, A. Scherer, F. McCormick, A. Wittinghofer, The 2.2 Å crystal structure of the ras-binding domain of the serine threonine kinase c-raf1 in complex with rap1a and a gtp analogue, Nature 375 (1995) 554–560.
[26] A.J. Scheidig, C. Burmester, R.S. Goody, The pre-hydrolysis state of p21(ras) in complex with GTP: new insights into the role of water molecules in the GTP hydrolysis reaction of ras-like proteins, Structure 7 (1999) 1311–1324.
[27] M.R. Ahmadian, C. Kiel, P. Stege, K. Scheffzek, Structural fingerprints of the Ras-GTPase activating proteins neurofibromin and p120GAP, J. Mol. Biol. 329 (2003) 699–710.