



ELSEVIER



# Evolutionary studies of ligand binding sites in proteins

Rafael J Najmanovich

Biological processes at their most fundamental molecular aspects are defined by molecular interactions with ligand–protein interactions in particular at the core of cellular functions such as metabolism and signalling. Divergent and convergent processes shape the evolution of ligand binding sites. The competition between similar ligands and binding sites across protein families create evolutionary pressures that affect the specificity and selectivity of interactions. This short review showcases recent studies of the evolution of ligand binding-sites and methods used to detect binding-site similarities.

## Address

Department of Pharmacology and Physiology, Faculty of Medicine, Université de Montreal, Montreal H3T 1J4, Quebec, Canada

Corresponding author: Najmanovich, Rafael J  
([rafael.najmanovich@umontreal.ca](mailto:rafael.najmanovich@umontreal.ca))

Current Opinion in Structural Biology 2017, 45:85–90

This review comes from a themed issue on **Engineering and design**

Edited by **Julia Shifman** and **Niv Papo**

<http://dx.doi.org/10.1016/j.sbi.2016.11.024>

0959-440/© 2016 Published by Elsevier Ltd.

## Introduction

The study of the evolution of ligand binding-sites in proteins, most notably enzymes has greatly benefited from our understanding of the evolution of protein structure [1–3] and function [4] as well as the availability of methods for the detection of binding-site similarities [5]. In recent years, insights on the evolution of protein function both as a result of divergent and convergent evolutionary processes have been the result of top-down and bottom-up approaches. Top-down approaches using the amassed available data in public repositories and efficient computational methods to perform large scale analyses clarified the relationship and extent of evolutionary connections between enzyme families whereas bottom-up approaches, particularly experimental approaches using directed evolution as a probing tool permitted to probe functional changes and evolutionary pathways in great detail. Recent breakthroughs include the understanding of the evolution of catalytic mechanisms across enzyme superfamilies and the appreciation of the role of convergent evolution. This short review focuses on small-molecule binding-sites but many of the

ideas may be applicable to the evolution of protein–protein interactions [6].

## Methods for the detection of binding-site similarities

Over the years a large number of methods to detect binding-site similarities have been developed. Only a few are presented here as representatives of the approaches mentioned. Different methods can be categorized according to the level of detail used to represent binding-sites, the methodology used to search for similarities and the scoring scheme (Table 1).

The detection of binding-site similarities depends on the capacity to detect or define binding-sites. Research on the detection of binding-sites aims in general at identifying cavities that are either biologically relevant (e.g., known to bind a ligand or allosterically affecting binding) or ‘druggable’. A number of methods exist for such a purpose, from purely geometric such as PASS [7], SURFNET [8] and its modern implementation within the NRGsuite PyMOL plugin [9] to methodologies that consider additional information such as evolutionary conservation [10] or energetic considerations as exemplified by PocketFinder [11]. The resulting cavities detected with any of the methods above (among others) can be used as input for the detection of similarities.

*Representation.* At a most basic or reduced level of representation, one can map binding-site residues onto the primary sequence and use pairwise or multiple sequence alignments to define a Tanimoto coefficient of binding-site sequence identity [12]. In other words, one can count the number of identical aligned binding-site residues  $c$  and normalize that number by the number of residues in either binding-site ( $a$  and  $b$  respectively) to obtain a Tanimoto score ( $c/(a + b - c)$ ). The eMatchSite algorithm goes a step further and creates sequence order-independent alignments of ligand binding-sites [13]. Considering the Functionalist principle discussed below, different methods representing binding-sites at increasing levels of biological complexity aim at detecting similarities that capture the biological information responsible for higher levels of conservation. Climbing the representation complexity ladder, at the level of structure we find a number of approaches to represent binding-sites, from C-alpha atoms and microenvironments to all-atom representations. PSILO [14], SOIPPA [15] as well as the C-alpha mode of IsoCleft [16,17] represent binding-sites via C-alpha atoms. APoc represents binding-sites utilizing the C-alpha and C-beta atoms as well as a classification of amino-acids into 8 classes [18]. Pre-defined atomic

Table 1

## A compilation of methods for the detection of binding-site similarities

| Method                          | Representation                    | Search               | Scoring          |
|---------------------------------|-----------------------------------|----------------------|------------------|
| eMatchSite <sup>a</sup> [13]    | Sequence                          | Geometric hashing    | Correlation      |
| PSILO <sup>b</sup> [14]         | C-alpha atoms                     | Exhaustive           | RMSD             |
| SOIPPA <sup>c</sup> [15]        | C-alpha atoms                     | Graph matching       | Profile distance |
| IsoCleft <sup>d</sup> [16,17]   | All-atoms                         | Graph matching       | Tanimoto/volume  |
| APoc <sup>e</sup> [18]          | C-alpha C-beta atoms and sequence | Structural alignment | PS-score         |
| CavBase <sup>f</sup> [19]       | Pseudo-centres                    | Graph matching       | Surface overlap  |
| SiteEngine <sup>g</sup> [20]    | Pseudo-centres                    | Geometric hashing    | Surface overlap  |
| PocketFEATURE <sup>h</sup> [21] | Microenvironments                 | Exhaustive           | Tanimoto score   |
| GRID-FLAP <sup>i</sup> [22]     | MIFs                              | Exhaustive           | Volume overlap   |
| IsoMIF [23*,24]                 | MIFs                              | Graph matching       | Tanimoto/volume  |

<sup>a</sup> Free source code and web-accessible at <http://www.brylinski.org/ematchsite>.

<sup>b</sup> Commercially available, Chemical Computing Group.

<sup>c</sup> Source code available as part of SMAP at [www.compsci.hunter.cuny.edu/~leixie/smap/smap](http://www.compsci.hunter.cuny.edu/~leixie/smap/smap) and web interface: [www.bioinfo.cs.pu.edu.tw/cloud-PLBS](http://www.bioinfo.cs.pu.edu.tw/cloud-PLBS).

<sup>d</sup> Free source code and web-accessible at [www.bcb.med.usherbrooke.ca/icfi](http://www.bcb.med.usherbrooke.ca/icfi).

<sup>e</sup> Source code freely available at <http://cssb.biology.gatech.edu/APoc>.

<sup>f</sup> Unknown availability.

<sup>g</sup> Free source code for non-commercial users and web-accessible at [www.bioinfo3d.cs.tau.ac.il/SiteEngine](http://www.bioinfo3d.cs.tau.ac.il/SiteEngine).

<sup>h</sup> Free source available at [www.simtk.org/projects/pocketfeature](http://www.simtk.org/projects/pocketfeature).

<sup>i</sup> Commercially available, Molecular Discovery.

<sup>j</sup> Free source code and web-accessible at [www.bcb.med.usherbrooke.ca/isomif](http://www.bcb.med.usherbrooke.ca/isomif).

pseudo-centres aim at capturing the presence of important interacting groups while decreasing the number of objects that need to be compared. Representative methods of this approach are CavBase [19] and SiteEngine [20]. PocketFEATURE [21] defines microenvironments at specific geometric centres of residues and calculated physico-chemical properties associated with the atoms present in concentric shells at different radii. The all-atom mode of IsoCleft [16,17] uses all non-hydrogen atoms to represent binding-sites. Further still in the complexity-representation ladder, we find methods that represent binding-sites by the potential interactions that could be made with particular chemical probes at different positions within the volume of the cavity using potential energy functions to define molecular interaction fields (MIFs). Notable methods in this category are GRID-FLAP [22] and IsoMIF [23\*,24]. The potential advantage of using MIFs rather than the specific positions of atoms or associated properties at the molecular surface is to account for different binding-site residue configurations that do not affect binding or cases where small differences can have drastic effects.

*Search and scoring.* In addition to representation, the other two pillars of any optimization problem are the method used for searching for solutions (in this case similarities between the binding-sites) and the scoring scheme. Predominantly used search algorithms are geometric hashing [25], graph matching [26] and exhaustive enumeration. The choice of method and its implementation depends on the type of representation used as that generally dictates the size of the search space. Different distance measures can be used to quantify similarity such as the

Tanimoto score, root mean square distance (RMSD) of the identified similarities after superimposition, and surface overlap (Table 1).

*Performance.* Despite the same purpose of detecting binding-site similarities, the different methods in Table 1 were developed with slightly different applications in mind and therefore were evaluated using tailor-made datasets. It would be interesting to add to this list of 'benchmark' datasets the Shoichet dataset discussed above [27\*\*]. As reported in [23\*], different methods perform well on particular datasets but poorly on others, with eMatchSite and IsoMIF having the largest average Area Under the receiver-operator Curve (AUC) across datasets at around 0.80. It is interesting to note that the two methods at the extremes of the scale of biological complexity representation discussed above have the best performance. However, unlike eMatchSite, IsoMIF displays a very low AUC variance, thus its performance is more robust across datasets. The wide variance in performance across datasets suggests that using multiple datasets is beneficial. Thus, the lack of any single ultimate benchmark dataset is a situation that should be maintained. Instead of a single benchmark dataset, even more benchmarking datasets should be used as part of a benchmark dataset pool. Whereas the advantages of such an approach are clear within the realm of methods for the detection of binding-site similarities, another field that would drastically benefit from such an approach is that of small-molecule docking simulation methods where benchmarking is dominated by the use of the Astex datasets [28,29] but different methods clearly vary in their performance when tested on different datasets [30].

## Divergent evolution

All members of a protein family, within a given organism (paralogs) and across organisms (orthologs) are the product of divergent evolution in response to distinct evolutionary forces. Within enzyme families in particular, the catalytic mechanism and reaction chemistry may be conserved but there may be changes in substrate specificity, in particular exploiting substrate promiscuity. In the promiscuous human cytosolic sulfotransferase family, structural binding-site similarities were related to substrate preferences among members [12,31], showing that better correlation with substrate preferences exists between binding-site similarities than global sequence similarities.

Promiscuity may also lead to changes in catalytic mechanism or reaction chemistry. This issue was studied by directed evolution approaches within a number of families belonging to the metallo-beta-lactamase superfamily [32]. In this superfamily catalytic promiscuity can be achieved or enhanced by a small number of mutations, linking different families. The work also suggests that selected changes are highly contingent on evolutionary history. Specifically, the authors observe many mutations having little or no effect on the wild type but becoming beneficial when occurring later on as part of a sequence of consecutive mutations acquired in subsequent steps of directed evolution [33]. The authors also observe diminishing returns for new mutations as a plateau is reached with respect to the contribution of additional mutations to the new catalytic activity. Furthermore, several mutations do not take place in the binding-site but among second-shell residues that help stabilize particular conformations of binding-site residues required for the change of functionality [34].

The Thornton group utilized existing data and integrated various computational methods to study 379 enzyme superfamilies linking changes in function to changes in reaction chemistry [35]. The authors observe examples of superfamilies with different reactions using the same catalytic machinery as well as cases where similar functions are performed by either different or similar residues. In fact, nearly 90% of all enzymatic functions in any one family are linked to functions in others and nearly 30% arose independently in two or more superfamilies unrelated by fold similarities. In the case of Flavin-dependent monooxygenases, the evolution of cofactor binding is intertwined with that of domain architecture, suggesting that cofactor binding may be a key constraint in the evolution of function within cofactor-dependent enzyme families [36].

## Convergent evolution

The PDB databank contains numerous examples of proteins with different fold classes [37,38] that evolved to bind the same molecule. Within our current understanding of protein evolution, such proteins are evolutionarily unrelated yet as they bind the same ligand, it has to be

assumed that this property evolved independently. This is particularly true for the evolution of cofactor binding-sites [39–44]. However, not just binding-sites of cofactors evolve convergently as discussed next.

Recently, the Shoichet group compiled a dataset of 62 pairs of protein–ligand complexes (59 different ligands) not including cofactors or small fragments, where within each pair the same ligand is bound to two proteins unrelated by fold [27]. The authors classify their entries into three distinct classes: (A) Ligand functional groups that interact with similar groups in the two protein binding sites (19 pairs); (B) A given ligand functional group that interacts with distinct protein groups in the two binding-sites (29 pairs); and (C) Different ligand functional groups that interact with similar or distinct binding-site functional groups (14 pairs). The authors compared the local environments surrounding ligand atoms using molecular potentials and found increasing levels of calculated electrostatic energy mean square differences between members of pairs across the A, B and C classes above. In two thirds of pairs, different multiple patterns of residues can recognize identical ligands. Interestingly, the authors suggest through statistical modelling that each ligand may have between two and five different recognition patterns, that is, binding-sites with different amino-acid patterns in proteins unrelated by fold.

As the Shoichet study shows, the same ligand, even the same functional groups in a ligand may be recognized by different functional groups in a protein. Thus, there will always be cases where recognizing that a given ligand binds to a sufficiently different binding-site is difficult or impossible given that a completely different set of interactions is exploited to stabilize the ligand. Therefore, whereas methods for the detection of binding site similarities are useful to study the evolution of binding-sites, there are limitations imposed by the nature of molecular recognition events.

The role of evolutionarily conserved residues in recognizing the same ligand across families was measured through the detection of binding-site similarities [16]. It has been observed that conserved residues are not good predictors of ligand binding, suggesting that evolutionarily conserved binding-site residues across protein folds are different. Indeed, a large dataset of neutral mutations can be compiled (Chénard T and Najmanovich R, unpublished results)<sup>1</sup> from the PDB database [45,46]. In many cases, one or more mutations of highly conserved binding-site residues in contact with a ligand do not prevent binding. As the evolutionary conservation of binding-site residues in contact with a ligand implies a functional role, it has to be remembered that molecular recognition is not a process that occurs in isolation but in

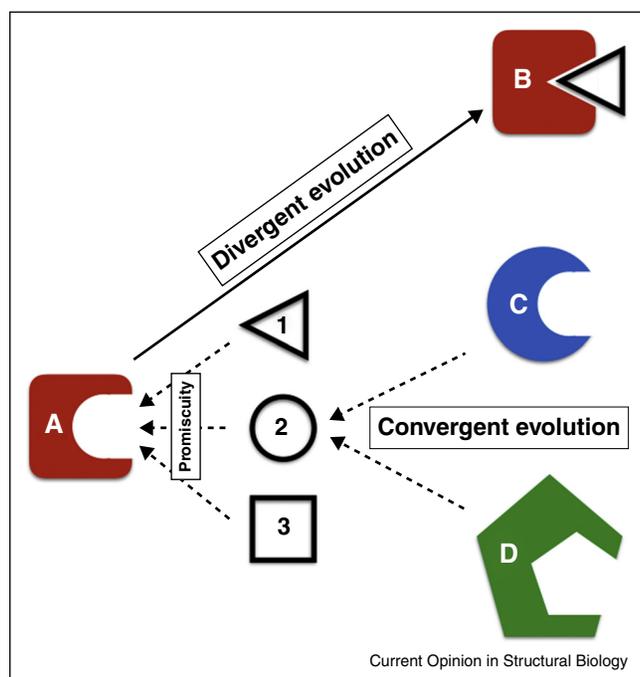
<sup>1</sup> <http://bcb.med.usherbrooke.ca/neutralmutdb.php>.

the presence of other ligands and proteins. Thus, evolutionarily conserved binding-site residues may be important not only for binding to a ligand but also for preventing binding of other molecules, that is, affecting both selectivity and specificity. As the cellular environment varies across cell types and cell cycle, the challenging task [47] of integrating structural and systems biology is required to fully understand molecular recognition.

The Skolnick group developed a library of 25 180 compact structures with random sequences selected to optimize thermodynamic stability based on a set of 1259 random PDB structures called the ART library [48]. The authors use the well established TM-score [49] (a measure of sequence independent structural similarity) to assess how closely ART library structures resemble the templates used to generate them. The authors note that despite being based on real structures, the average TM-score of artificial (ART) structures with respect to their original templates is 0.33, close to the value of

0.30 obtained for random structural alignments. The authors proceeded to use the PS-score [48] (a measure of binding-site similarity that accounts for C-alpha distances, and C-alpha/C-beta orientations and 8 amino-acid classes) to detect binding-site similarities among ART pockets, those in a dataset of 5371 non-redundant PDB structures and between the ART and non-redundant PDB datasets. The authors find that a small number of pockets, either natural or artificial is enough to detect similarities in all pockets. Based on this, it is suggested that space of binding-site pockets is likely complete and arise as a consequence of imperfections in the geometric packing of proteins. Furthermore, the authors demonstrate that the convergent evolution of binding-sites, rather than being a rare evolutionary event is instead unavoidable. The authors paint a gloomy picture for drug design with respect to the consequences of promiscuity and drug cross-reactivity but also suggest that this increases the potential for repurposing drugs [50<sup>••</sup>]. However, small-differences in binding-sites can have drastic effects with respect to binding [16,23<sup>•</sup>]. This gives plenty of room to maneuver in modulating the selectivity and specificity of drugs by means of small alterations in their chemical composition to exploit small differences between otherwise highly similar binding-sites. The different processes that affect the evolution of ligand binding-sites in proteins are summarised in Figure 1.

Figure 1



Different processes shape the functional evolution of ligand binding-sites in proteins. A protein A may bind different ligands or catalyse different reactions (both cases represented by shapes 1, 2 and 3). Mutations in the binding-site or elsewhere may lead to fixing one particular function B). When comparing the mutations that lead A to B only with respect to ligand/reaction 1, these mutations will be considered as neutral. However, they affect the binding of other ligands. Thus the need to consider molecular recognition within the larger context of cellular function. Convergent evolutionary processes may lead to different proteins (C and D), evolutionarily unrelated to A, to bind the same ligand as A or perform the same reaction. The same set (C) or a different set (D) of interactions can evolve to bind the same ligand or perform the same reaction. Just as mutations may be 'neutral', they may also cause drastic effects.

### The functionalist principle in biology

The functionalist principle in biology<sup>2</sup> describes the observation that sequence is less conserved than structure that itself is less conserved than function. The latter defined here as the capacity to bind the same ligand or in the case of enzymes as having the same reaction chemistry. In cases of convergent evolution, both sequence and structural similarities could be minimal but the pair of proteins may have the same function. Cases of divergent evolution are counter-examples of the functionalist principle where, as a result of a small number of mutations either or both, function and structure may change. At an extreme, there are cases such as for example PrP<sup>c</sup> and PrP<sup>sc</sup>, two distinct structural forms of the same prion protein where one is pathogenic and the other is not [51]. This may not be an outlier as naturally unstructured proteins can in principle undergo different conformational selection pathways in different situations [52,53]. Comparison of TM-scores and PS-scores in the ART library led Skolnick & Gao to suggest that 'protein pocket structure is likely strongly driven by selection for fold stability rather than by function' [48]. However, it is important to keep in mind that structural conservation

<sup>2</sup> The term 'Functionalist principle' was suggested by Dr. Simone Botti after the architect Louis Sullivan who coined the phrase 'Form follows function' from which the concept of Functionalism in architecture was developed. Clearly, opposed to biology, structure in architecture is designed under Functionalism whereas in biology function is selected through natural selection.

is loosely defined as even a single amino acid change will lead to small differences in the structure, dynamics and function [54–56] of proteins that are glossed over when discussing examples of the functionalist principle. Ultimately, the use of the term has to be understood as grouping a set of observations that are only true in broad terms instead of as a rule.

## Conclusions

Recent developments in our comprehension of binding-site evolution involve clarification of connections between functionally distinct protein families, demonstrating the fact that small differences can have drastic effects and the role of promiscuity taking advantage of binding-site residues that are non-specificity-determining or involved in selectivity in facilitating functional shifts.

## Acknowledgements

The author would like to apologize to those cited in this review in case of any misrepresentation of their work as well as to the authors of the numerous outstanding papers that for lack of space had to be omitted.

The author would like to acknowledge the contributions of Dr. Simone Botti, Dr. Harel Weinstein, Dr. Ilan Samish, Dr. Burkhard Rost, Dr. Hernan Chaimovich, Dr. Itamar Borukhov and Dr. Matthew Bashton in discussions in the Facebook social media platform regarding the Functionalist principle in biology.

RJN is part of the Québec network for research on protein function, structure and engineering (PROTEO) and Groupe de Recherche Axé sur la Structure des Protéines (GRASP).

**Funding:** RJN is the recipient of a Junior II salary fellowship from the Fonds de Recherche du Québec–Santé (FRQ-S). This project was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) [Discovery Grant RGPIN-2014-05766].

## References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Goldstein RA: **The structure of protein evolution and the evolution of protein structure.** *Curr Opin Struct Biol* 2008, **18**:170-177.
  2. Valas RE, Yang S, Bourne PE: **Nothing about protein structure classification makes sense except in the light of evolution.** *Curr Opin Struct Biol* 2009, **19**:329-334.
  3. Xia Y, Levitt M: **Simulating protein evolution in sequence and structure space.** *Curr Opin Struct Biol* 2004, **14**:202-207.
  4. Martínez Cuesta S, Rahman SA, Furnham N, Thornton JM: **The classification and evolution of enzyme function.** *Biophys J* 2015 <http://dx.doi.org/10.1016/j.bpj.2015.04.020>.
  5. Najmanovich RJ, Torrance JW, Thornton JM: **Prediction of protein function from structure: insights from methods for the detection of local structural similarities.** *BioTechniques* 2005, **38**:847, 849, 851.
  6. Marsh JA, Teichmann SA: **Structure, dynamics, assembly, and evolution of protein complexes.** *Annu Rev Biochem* 2015, **84**:551-575.
  7. Brady GP, Stouten PF: **Fast prediction and visualization of protein binding pockets with PASS.** *J Comput Aided Mol Des* 2000, **14**:383-401.
  8. Laskowski R: **Surfnet – a program for visualizing molecular-surfaces, cavities, and intermolecular interactions.** *J Mol Graph* 1995, **13**:323-330.
  9. Gaudreault F, Morency L-P, Najmanovich RJ: **NRGsuite: a PyMOL plugin to perform docking simulations in real time using FlexAID.** *Bioinformatics* 2015 <http://dx.doi.org/10.1093/bioinformatics/btv458>.
  10. Glaser F, Morris RJ, Najmanovich RJ, Laskowski RA, Thornton JM: **A method for localizing ligand binding pockets in protein structures.** *Proteins* 2006, **62**:479-488.
  11. An J, Totrov M, Abagyan R: **Pocketome via comprehensive identification and classification of ligand binding envelopes.** *Mol Cell Proteom* 2005, **4**:752-761.
  12. Najmanovich RJ, Allali-Hassani A, Morris RJ, Dombrovsky L, Pan PW, Vedadi M, Plotnikov AN, Edwards AM, Arrowsmith C, Thornton JM: **Analysis of binding site similarity, small-molecule similarity and experimental binding profiles in the human cytosolic sulfotransferase family.** *Bioinformatics* 2007, **23**:e104-e109.
  13. Brylinski M: **eMatchSite: sequence order-independent structure alignments of ligand binding pockets in protein models.** *PLoS Comput Biol* 2014, **10**:e1003829.
  14. Feldman HJ, Labute P: **Pocket similarity: are  $\alpha$  carbons enough?** *J Chem Inf Model* 2010, **50**:1466-1475.
  15. Xie L, Bourne PE: **Detecting evolutionary relationships across existing fold space, using sequence order-independent profile-profile alignments.** *Proc Natl Acad Sci U S A* 2008, **105**:5441-5446.
  16. Najmanovich R, Kurbatova N, Thornton J: **Detection of 3D atomic similarities and their use in the discrimination of small molecule protein-binding sites.** *Bioinformatics* 2008, **24**:i105-i111.
  17. Kurbatova N, Chartier M, Zylber MI, Najmanovich RJ: **IsoCleft Finder – a web-based tool for the detection and analysis of protein binding-site geometric and chemical similarities.** *F1000Res* 2013, **2**:117.
  18. Gao M, Skolnick J: **APoc: large-scale identification of similar protein pockets.** *Bioinformatics* 2013, **29**:597-604.
  19. Schmitt S, Kuhn D, Klebe G: **A new method to detect related function among proteins independent of sequence and fold homology.** *J Mol Biol* 2002, **323**:387-406.
  20. Shulman-Peleg A, Nussinov R, Wolfson H: **Recognition of functional sites in protein structures.** *J Mol Biol* 2004, **339**:607-633.
  21. Liu T, Altman RB: **Using multiple microenvironments to find similar ligand-binding sites: application to kinase inhibitor binding.** *PLoS Comput Biol* 2011, **7**:e1002326.
  22. Sciabola S, Stanton RV, Mills JE, Flocco MM, Baroni M, Cruciani G, Perruccio F, Mason JS: **High-throughput virtual screening of proteins using grid molecular interaction fields.** *J Chem Inf Model* 2010, **50**:155-169.
  23. Chartier M, Najmanovich R: **Detection of binding site molecular interaction field similarities.** *J Chem Inf Model* 2015, **55**:1600-1615.
- The authors present the IsoMIF software and compare its performance to a number of existing methods.
24. Chartier M, Adriansen E, Najmanovich R: **IsoMIF Finder: online detection of binding site molecular interaction field similarities.** *Bioinformatics* 2015, **32**:btv616-btv623.
  25. Rigoutsos I, Wolfson H: **Geometric hashing.** *IEEE Comput Sci Eng* 1997, **4**:9-9.
  26. Gross J, Yellen J, Rosen KH: *Handbook of Graph Theory*. 2004.
  27. Barelier S, Sterling T, O'Meara MJ, Shoichet BK: **The recognition of identical ligands by unrelated proteins.** *ACS Chem Biol* 2015 <http://dx.doi.org/10.1021/acscchembio.5b00683>.
- Development of the dataset of non-homologous proteins that bind identical ligands and the definition of different molecular recognition classes.

28. Verdonk ML, Mortenson PN, Hall RJ, Hartshorn MJ, Murray CW: **Protein–ligand docking against non-native protein conformers**. *J Chem Inf Model* 2008, **48**:2214–2225.
29. Hartshorn MJ, Verdonk ML, Chessari G, Brewerton SC, Mooij WTM, Mortenson PN, Murray CW: **Diverse, high-quality test set for the validation of protein–ligand docking performance**. *J Med Chem* 2007, **50**:726–741.
30. Gaudreault F, Najmanovich RJ: **FlexAID: revisiting docking on non-native-complex structures**. *J Chem Inf Model* 2015, **55**:1323–1336.
31. Allali-Hassani A, Pan PW, Dombrovski L, Najmanovich R, Tempel W, Dong A, Loppnau P, Martin F, Thornton J, Thonton J *et al.*: **Structural and chemical profiling of the human cytosolic sulfotransferases**. *PLoS Biol* 2007, **5**:e97.
32. Baier F, Tokuriki N: **Connectivity between catalytic landscapes of the metallo- $\beta$ -lactamase superfamily**. *J Mol Biol* 2014, **426**:2442–2456.
33. Miton CM, Tokuriki N: **How mutational epistasis impairs predictability in protein evolution and design**. *Protein Sci* 2016, **25**:1260–1272.
- The role of evolutionary history is explored using directed evolution.
34. Yang G, Hong N, Baier F, Jackson CJ, Tokuriki N: **Conformational tinkering drives evolution of a promiscuous activity through indirect mutational effects**. *Biochemistry* 2016, **55** acs.biochem.6b00561-4593.
- Utilizing directed evolution the authors demonstrate the role of second-shell residues in the evolution of enzymatic function between protein families.
35. Furnham N, Dawson NL, Rahman SA, Thornton JM, Orengo CA: **Large-scale analysis exploring evolution of catalytic machineries and mechanisms in enzyme superfamilies**. *J Mol Biol* 2016, **428**:253–267.
- Analysis across all enzyme superfamilies of the relationship between the evolution of structure and that of reaction chemistry.
36. Mascotti ML, Juri Ayub M, Furnham N, Thornton JM, Laskowski RA: **Chopping and changing: the evolution of the flavin-dependent monooxygenases**. *J Mol Biol* 2016, **428**:3131–3146.
37. Sillitoe I, Lewis TE, Cuff A, Das S, Ashford P, Dawson NL, Furnham N, Laskowski RA, Lee D, Lees JG *et al.*: **CATH: comprehensive structural and functional annotations for genome sequences**. *Nucleic Acids Res* 2015, **43**:D376–D381.
38. Andreeva A, Howorth D, Chothia C, Kulesha E, Murzin AG: **SCOP2 prototype: a new approach to protein structure mining**. *Nucleic Acids Res* 2014, **42**:D310–D314.
39. Holliday GL, Fischer JD, Mitchell JBO, Thornton JM: **Characterizing the complexity of enzymes on the basis of their mechanisms and structures with a bio-computational analysis**. *FEBS J* 2011, **278**:3835–3845.
40. Fischer JD, Holliday GL, Rahman SA, Thornton JM: **The structures and physicochemical properties of organic cofactors in biocatalysis**. *J Mol Biol* 2010, **403**:803–824.
41. Fischer JD, Holliday GL, Thornton JM: **The CoFactor database: organic cofactors in enzyme catalysis**. *Bioinformatics* 2010, **26**:2496–2497.
42. Kahraman A, Morris RJ, Laskowski RA, Favia AD, Thornton JM: **On the diversity of physicochemical environments experienced by identical ligands in binding pockets of unrelated proteins**. *Proteins* 2010, **78**:1120–1136.
43. Kahraman A, Morris RJ, Laskowski RA, Thornton JM: **Shape variation in protein binding pockets and their ligands**. *J Mol Biol* 2007, **368**:283–301.
44. Stegemann B, Klebe G: **Cofactor-binding sites in proteins of deviating sequence: comparative analysis and clustering in torsion angle, cavity, and fold space**. *Proteins* 2012, **80**:626–648.
45. Rose PW, Rose PW, Prli A, Pric A, Bi C, Bi C, Bluhm WF, Bluhm WF, Christie CH, Christie CH *et al.*: **The RCSB Protein Data Bank: views of structural biology for basic and applied research and education [Internet]**. *Nucleic Acids Res* 2015, **43**:D345–D356.
46. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE: **The PROTEIN DATA BANK**. *Nucleic Acids Res* 2000, **28**:235–242.
47. Samish I, Bourne PE, Najmanovich RJ: **Achievements and challenges in structural bioinformatics and computational biophysics**. *Bioinformatics* 2015, **31**:146–150.
48. Skolnick J, Gao M: **Interplay of physics and evolution in the likely origin of protein biochemical function**. *Proc Natl Acad Sci U S A* 2013, **110**:9344–9349.
49. Zhang Y, Skolnick J: **Scoring function for automated assessment of protein structure template quality**. *Proteins* 2004, **57**:702–710.
50. Skolnick J, Gao M, Roy A, Srinivasan B, Zhou H: **Implications of the small number of distinct ligand binding pockets in proteins for drug discovery, evolution and biochemical function**. *Bioorg Med Chem Lett* 2015, **25**:1163–1170.
- Discussion of the results obtained using the ART library of protein models with respect to the convergent evolution of binding pockets.
51. Riesner D: **Biochemistry and structure of PrP(C) and PrP(Sc)**. *Br Med Bull* 2003, **66**:21–33.
52. Wright PE, Dyson HJ: **Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm**. *J Mol Biol* 1999, **293**:321–331.
53. Oldfield CJ, Dunker AK: **Intrinsically disordered proteins and intrinsically disordered protein regions**. *Annu Rev Biochem* 2014, **83**:553–584.
54. Frappier V, Najmanovich RJ: **A coarse-grained elastic network atom contact model and its use in the simulation of protein dynamics and the prediction of the effect of mutations**. *PLoS Comput Biol* 2014, **10**:e1003569.
55. Frappier V, Chartier M, Najmanovich RJ: **ENCoM server: exploring protein conformational space and the effect of mutations on protein function and stability**. *Nucleic Acids Res* 2015, **43**:W395–W400.
56. Frappier V, Najmanovich RJ: **Vibrational entropy differences between mesophile and thermophile proteins and their use in protein engineering**. *Protein Sci* 2015, **24**:474–483.