Journal of Molecular Graphics and Modelling 84 (2018) 69–73

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Shape complementarity at protein interfaces via global docking optimisation

Gareth Williams

Bioinformatics, Wolfson CARD, The Wolfson Wing, Hodgkin Building, King's College London, SE1 1UL, UK

ARTICLE INFO

Article history:
Received 19 April 2018
Received in revised form
11 June 2018
Accepted 12 June 2018
Available online 15 June 2018

Keywords:
Protein structure
Protein-protein docking
Shape complementarity

ABSTRACT

Protein complexes are characterised by shape complementarity at the interface. Here we present a simple fast global shape fitting algorithm to investigate the extent to which interfaces are global minima of complementarity. The algorithm is applied to a varied set of hetero and homo complexes and complexes between complexes showing that over 90% of large interfaces are global maxima in the space of shape complementarity.

© 2018 Published by Elsevier Inc.

1. Introduction

Protein interaction interfaces are characterised by a high degree of shape complementarity [1,2]. Lawrence and Coleman [1] gave a quantitative score to the goodness of fit at a protein interface through a dot product of surface vectors corresponding to proximal atoms. This technique served as the basis for a concise Ramachandran-like 2D plot representation of protein interfaces combining shape and electrostatic complementarity (Sc and EC) [3]. Machine learning implementations based on Sc, EC and interface size have been successful in filtering out true native-like docked conformations from a dataset of possible poses [4]. The extent to which crystallographic complex solutions deviate from interface complementarity can also be scored and visualised via a small radius probe, implemented with the Molprobity web tool [5]. Molprobity has served as an effective tool for crystal structure optimisation.

Global shape complementarity docking algorithms such as GRAMM have shown that native-like docked complexes emerge as those with optimal surface overlap at various levels of coarse granularity [6]. Such exhaustive searches are over a six dimensional space with an additional scoring of the interface and are usually speeded up with techniques such as the fast Fourier transform (FFT). This methodology has been extended in FTDock [7] to include an electrostatic filter to separate high complementarity poses. The FFT overlap calculation speed up has been the basis of ClusPro [8], where high overlap conformations filtered based on statistical potentials and then clustered, and DOT [9]. Alternatively, protein structures have been approximated with a low order spherical polar Fourier expansion with a resulting relatively fast 'surface skin' correlation calculation [10]. Non-exhaustive techniques have also been developed based on defining surface features according to local convexity, concavity and flatness. In this PatchDock approach only conformations with matching patches are scored for complementarity [11]. In a further level of abstraction, the binding site shape has been shown to be describable with the first few terms of a Zernike 3D shape descriptor polynomial leading to a relatively rapid complementarity calculation [12]. The methodology presented here is based on an initial coarse grained exploration of conformations pivoted on surface atom surface point pairs followed by a fine scale analysis of limited set of putative binding conformations. The methodology recovers 92% of a mixed set of protein complex conformations.

2. Results

The algorithm was tested on a mixed set of 278 protein complexes. These consisted of 87 homo-oligomer complexes [13] and 191 transient hetero-oligomer complexes [14], with 65 of these complexes involving more than two proteins. The complexes are non-redundant in the sense that they don't share sequence homology at the interface. The results described below are not sensitive to the particular set of complexes examined. The first stage

E-mail address: gareth.2.williams@kcl.ac.uk.

https://doi.org/10.1016/j.jmgm.2018.06.011
1093-3263/© 2018 Published by Elsevier Inc.

coarse grained docking is sufficient to identify 61% of the complexes as optimal overlap conformations. A native-like docked conformation is called optimal if it comes in the top four ranked conformations according to the given scoring system. This percentage rises to 81% with the fine grained optimisation. The likelihood of a binding conformation being an optimal in shape complementarity increases with the size of the binding interface, see Table 1. In particular, for complexes with a number of atom contacts (4 Å proximity) greater than 400 92% are complementarity optimal. This constitutes 80% of the complexes. It appears that measuring the size of the interaction through the amount of buried accessible surface area (ASA) is worse at segregating optimal from non-optimal complementarity, see Fig. 1. Here, only 84% of the top 80% ASA complexes are complementarity optimal. An example docking run is shown in Fig. 2. Here, an antibody light/heavy chain pair is docked with its target (von Willebrand factor pdb accession 1fe8). The 'ligand' scores highly at multiple sites on the antibody, with two conformations (ranked 1 and 3) in the top five aligning with the native structure.

The atom type content of the interface is invisible to the analysis so far. A simple way to introduce atom type content in interface description is through a vector over a relatively small set of properties. In particular, five types of atom are considered: neutral, donor, acceptor, positive, negative. Thus each docking pose is associated with a docking matrix. A simple linear model can then be used to maximise the score associated with the native-like docking conformation. In particular, collecting 190 poses with corresponding RMSDs a linear model predicts slightly more native-like docked conformations, 86% and 96% of large interfaces. The relative contribution of the various atom pairings to the native-like docked interface relative to other complementary interfaces is shown in Table 2. The main contribution by virtue of being the dominant atop type comes from neutral pairings. As expected, opposite charge pairings and pairings between acceptor/donor and positive/negative atoms also contribute positively. The biggest effect is on the smaller interfaces where the native-like docked conformations now constitute 49% of the high scoring complementarity poses as opposed to 40% without atom type information.

Amino acid preferences in both inter- and intra-protein interaction have been the subject of much research. The propensities for internal contacts have been developed into statistical potentials [15–17] that have been employed in protein folding simulations [18]. Propensities of amino acid types at protein interfaces have been effectively deployed as supplements to docking scores [19–23]. With this in mind it is of interest to investigate to what extent high complementarity interfaces segregate between native and non-native on the basis of amino acid content. Amino acid type data can be introduced in the same way as atom type data. However, a linear model fit based on the amino acid content and contact number at the interface only results in a moderate improvement in predictability from 81% to 83%. The beta factors for the amino acid contribution have a small but significant correlation with the probabilities associated with the Miyazawa Jernigan amino acid interaction energies (Pearson correlation −0.25 Zscore 3.69).

Table 1
The protein complexes used in the analysis. The protein data bank identifier is given together with the two chain identifiers. The colon separates the single or multiple domains to be docked.

12asB:A	1bkd R:S	1dfj l:E	1fbi HL:X	1i7w A:B	1ncc N:HL	1ugh I:E
1a2k A:D	1bkp B:A	1dhk A:B	1fc2 C:D	1i8l A:C	1nfd EF:AB	1vfr B:A
1a2y AB:C	1blx A:B	1dkg AB:D	1fe8 A:HL	1iar A:B	1nrn HL:R	1vok B:A
1a4i B:A	1bmd B:A	1dor B:A	1fip B:A	1ib1 AB:E	1nse B:A	1vrkA:B
1a4u B:A	1brml A:C	1dpg B:A	1fj1 AB:F	1ibr A:B	1nsn HL:S	1wej HL:F
1a4yA:B	1bp3 A:B	1dpj A:B	1fle E:I	1ief AB:I	1nsy B:A	1wql G:R
1aa7 B:A	1bqq T:M	1dqj AB:C	1flt VW:Y	1icw B:A	1osp HL:0	1wtl B:A
1acb E:l	1brw B:A	1dqs B:A	1fns HL:A	1ihs HL:I	1pgt B:A	1www VW:Y
1ad3 B:A	1bsl B:A	1ds6 A:B	1foe A:B	1im3 A:D	1ppf l:E	1xso B:A
1ade B:A	1bsr B:A	1dtd A:B	1fq1 A:B	1im9 A:D	1pre B:A	1ycs A:B
1adq A:HL	1buh A:B	1du3 DF:A	1fqk A:B	1iinb B:A	1qa9 A:B	1zbd A:B
1afw B:A	1bvn P:T	1dx5 AM:I	1fqv A:B	1iqd AB:C	1qavA:B	2arc B:A
1agr A:E	1bxg B:A	1dxg B:A	1fro E:A	1ira X:Y	1qfh B:A	2btc l:E
1aip A:CD	1bxk B:A	1dzb X:A	1fsk A:BC	1isa B:A	1qfu AB:HL	2btfP:A
1ajs B:A	1bzq A:L	1eOo A:B	1fyh A:B	litb A:B	1qhi B:A	2ccy B:A
1ak4 A:D	1cly A:B	1e44 A:B	1g3n A:C	1ivy B:A	1qkz A:HL	2bdh B:A
1aq6 B:A	1c4z A:D	1e6j HL:P	1g4u R:S	1j7vR:L	1qmz A:B	2hmi B:CD
1atn A:D	1caO BC:D	1e96 A:E	1g4y R:B	1jdp H:A	1qoO A:DE	2jel HL:P
1auo B:A	1cd9 A:B	1eai A:C	1g73 A:D	1jhl HL:A	1qo3 A:CD	2lig B:A
1ava A:C	1cd9 A:B	1eay A:C	1g9m C:G	1jiw P:l	1qr2 B:A	2mcg 2:1
1avg HL:I	1cdk A:l	1ebh B:A	1g9m G:HL	1jlt A:B	1r2f B:A	2nac B:A
1avw A:B	1cdm A:B	1ebp A:CD	1gcq B:C	1jma A:B	1reg Y:X	2ohx B:A
1avz B:C	1cg2 C:E	1efu A:B	1gh6 A:B	1jps HL:T	1rfb B:A	2pcc A:B
1axi A:B	1chm B:A	1egj A:HL	1glO l:E	1jrh HL:I	1rlb ABCD:F	2sicE:I
1ay7 A:B	1cho E:F	1eja A:B	1gl4 A:B	1jtd A:B	1rrp A:B	2spc B:A
1azz A:CD	1cic AB:CD	1emv A:B	1got A:B	1jtgAiB	1sbb A:B	2vir AB:C
1b2s A:D	1clv l:A	1eo8 AB:HL	1hcf AB:Y	1jtp A:L	1ses B:A	3bth l:E
1b3a B:A	1cmb B:A	1es7 AC:D	1he1 A:C	1k4c AB:C	1sgp l:E	3dap E:A
1b5e B:A	1cmx A:B	1euv A:B	1hez AB:E	1k90 A:D	1slt B:A	3sdh B:A
1b67B:A	1cn4 AB:C	1ev2 A:E	1hia AB:l	1k9o E:l	1slu A:B	3ygsP:C
1b6c A:B	1cnz B:A	1f02I:T	1hjr D:B	1kac A:B	1smn B:A	4chaC:B
1b8a B:A	1coz B:A	1f34 A:B	1hss E:A	1kba B:A	1smt B:A	4htc HL:I
1b8j B:A	lcse l:E	1f3v A:B	1hx1 A:B	1kcg AB:C	1sox B:A	4kbp C:B
1bbh B:A	1cxz A:E	lf51 AE:E	1hxp E:A	1kig HL:I	1stf l:E	4sgb l:E
1bdO B:A	1d2z A:B	1f5q A:E	1hyr AB:C	1lfd AC:B	ltbr HL:R	5rub B:A
1bdj A:B	1d5m A:C	lf60 A:B	1i1r A:B	1lpb A:E	1tcl B:A	7cei A:B
1biq B:A	ld6r l:A	1f7z A:l	1i2m A:B	1lyn B:A	1tmq A:B	8prk E:A
1bis B:A	1daa B:A	1f93 AE:EF	1i4d AB:D	1m6p B:A	1tnr A:R	9wga B:A
1bjw HL:W	1dee CD:G	1fak HL:T	1i4o A:C	1mkb B:A	1toc R:AB	
1bjw B:A	1dev A:B	1fak l:HL	1i5k A:C	1mlc AB:E	1tx4 A:B	

Download English Version:

https://daneshyari.com/en/article/6877337

Download Persian Version:

https://daneshyari.com/article/6877337

Daneshyari.com