



UNIVERSIDADE
CATOLICA
PORTUGUESA

INSTITUTO DE
CIÊNCIAS DA SAÚDE

INTERATÓMICA DA CAVIDADE ORAL – *ORALINT V2.0*

Dissertação apresentada à Universidade Católica Portuguesa para
obtenção do grau de Mestre em Medicina Dentária

Por:

Sara Filipa Nunes Lopes

Viseu 2015



UNIVERSIDADE
CATOLICA
PORTUGUESA

INSTITUTO DE
CIÊNCIAS DA SAÚDE

INTERATÓMICA DA CAVIDADE ORAL – *ORALINT V2.0*

Dissertação apresentada à Universidade Católica Portuguesa para obtenção
do grau de Mestre em Medicina Dentária

Sob a orientação da Professora Doutora Maria José Correia e a
coorientação do Professor Doutor Joel Arrais

Por:

Sara Filipa Nunes Lopes

Viseu 2015

“Eles não sabem, nem sonham, que o sonho comanda a vida,
que sempre que um homem sonha o mundo pula e avança
Como bola colorida entre as mãos de uma criança”

António Gedeão

Índice

Agradecimentos	ix
Lista de Abreviaturas	xi
Resumo	xiii
Abstract	xv
Nota Prévia	xvii
1-INTRODUÇÃO	1
1.1 Microbioma Humano	5
1.2 Cavidade Oral	5
1.3 Ecossistema Oral.....	6
1.3.1 Microbioma Oral	6
1.4 Biofilme Oral	6
1.5 Interatômica	8
1.5.1 Interação proteína-proteína (PPIs).....	8
1.5.2 Interatoma microbioma- hospedeiro	9
1.6 Ferramentas bioinformáticas para o estudo do interatoma da cavidade oral 10	
1.6.1 OralOme, Oralcard e OralInt.....	11
1.6.2 Cytoscape e os seus layouts como software informático para análise das redes de PPIs.....	12
1.7 Análise de redes de PPIs.....	12
2-OBJETIVOS	15
3-MATERIAS E MÉTODOS	19
3.1 Proteoma Oral.....	21
3.2 Análise das redes geradas.....	22
3.3. Utilização do Cytoscape para análise da rede.....	24
3.4 Criação de sub-redes para interpretação biológica da rede do interactoma oral.....	25
3.5 Ferramentas Bioinformáticas para recolha de informação funcional e biológica das proteínas	27
4- RESULTADOS E DISCUSSÃO	29
4.1 Interactoma da cavidade oral	31

4.1.1	Proteínas usadas pelo <i>OralInt v2.0</i>	31
4.1.2	Proteínas presentes na rede do interactoma oral em estudo.....	32
4.1.3	Tipo de interações previstas pelo <i>OralInt v2.0</i>	34
4.1.4	Análise topológica da rede do interactoma oral.....	35
4.2	Interações previstas para proteínas de microrganismos associados a patologias orais.....	39
4.2.1	Interactoma oral de <i>Porphyromonas gingivalis</i>	42
4.2.2	Interactoma oral de <i>Streptococcus mutans</i>	52
5-	CONCLUSÃO	59
6-	BIBLIOGRAFIA.....	63

Índice de figuras

Figura 1- Interações previstas pelo <i>OralInt v2.0</i> importados para o <i>Cytoscape</i> , com a definição da fonte e alvo de interação, assim como a classificação do tipo de interação.	23
Figura 2- Atributos de cada proteína. Importação para o <i>Cytoscape</i> referentes à informação pertencente a cada proteína presente na rede.	24
Figura 3- Ferramentas que o <i>Cytoscape</i> disponibiliza para a análise topológica da rede.	24
Figura 4 - Página inicial do Uniprot obtida através do site http://www.uniprot.org/ para recolha de informação das proteínas. Exemplo: pesquisa da proteína Q99616 onde consta a informação referente à proteína em questão.....	27
Figura 5- Página inicial do BioGrid obtida através do site http://thebiogrid.org/ para consulta de interações de proteínas já descritas na literatura.	28
Figura 6 - Página inicial do Pubmed obtido do site http://www.ncbi.nlm.nih.gov/pubmed/	28
Figura 7- Número de proteínas microbianas distribuídas por género microbiano presentes na cavidade oral que o <i>OralInt v2.0</i> analisou.	31
Figura 8 - Número de proteínas de origem microbiana distribuídas por género microbiano presentes nas 200000 interações previstas pelo <i>OralInt v2.0</i> com <i>score</i> superior a 0,925.	33
Figura 9- Distribuição da percentagem dos tipos de interações estabelecidas entre as proteínas presentes nas 200000 interações selecionadas.	35
Figura 10- Representação da percentagem das proteínas agrupadas de acordo com o seu valor de <i>degree</i> na rede em estudo	36
Figura 11- Representação das interações previstas pelo <i>OralInt v2.0</i> entre os 4 grupos: as proteínas humanas (vermelho), proteínas de espécies microbianas associadas à cárie (verde), proteínas de espécies microbianas associadas à doença periodontal (laranja) e as restantes proteínas microbianas (azul). O tamanho da caixa é proporcional ao número de proteínas pertencentes a cada grupo. As interações entre as proteínas dos diferentes grupos são representadas pelas linhas de forma proporcional ao número de interações previstas pelo <i>OralInt v2.0</i>	41

Figura 12- Número de proteínas distribuídas por espécie que interagem com a proteína <i>gingipain R2</i> (P95493) previstas pelo <i>OralInt v2.0</i>	49
Figura 13- Representação das proteínas distribuídas por género que o interactoma oral do <i>Streptococcus mutans</i> apresenta como interações para a proteína <i>glucosyltransferase-I</i> (P08987).....	55
Figura 14- Representação das proteínas distribuídas por género que o interactoma oral do <i>Streptococcus mutans</i> apresenta como interações para a proteína <i>glucocosltransferase-S</i> (P49331).....	56

Índice de tabelas

Tabela 1- Códigos Uniprot correspondentes a proteínas <i>gingipains</i> produzidas pela espécie <i>Porphyromonas gingivalis</i>	25
Tabela 2- Códigos Uniprot correspondentes a proteínas <i>FimA</i> produzidas pela espécie <i>Porphyromonas gingivalis</i>	26
Tabela 3- Códigos Uniprot da proteína <i>glucosyltransferase</i> produzidas pela espécie <i>Streptococcus mutans</i>	26
Tabela 4- Proteínas <i>hub</i> da rede do interactoma oral.	37
Tabela 5- Espécies microbianas presentes na amostra do <i>OralInt v2.0</i> associadas às duas patologias orais com maior impacto na medicina dentária (cárie e doença periodontal).....	40
Tabela 6- Proteínas <i>hub</i> da rede do interactoma oral de <i>Porphyromonas gingivalis</i>	43
Tabela 7- Proteínas de <i>Porphyromonas gingivalis</i> com maior valor de <i>degree</i> na rede do interactoma oral de <i>Porphyromonas gingivalis</i>	45
Tabela 8- Relevância das interações estabelecidas pela proteína <i>Lys-gingipain</i> (B2RLK2) produzida pela <i>Porphyromonas gingivalis</i>	47
Tabela 9- Relevância das interações estabelecidas pela proteína <i>Lys-gingipain</i> (P72197) produzida pela <i>Porphyromonas gingivalis</i>	47
Tabela 10- Relevância das interações estabelecidas pela proteína <i>Lys-gingipain 381</i> (P72194) produzida pela <i>Porphyromonas gingivalis</i>	48
Tabela 11- Relevância das interações estabelecidas pela proteína <i>Major fimbrial subunit protein</i> (P59914) produzida pela <i>Porphyromonas gingivalis</i>	51
Tabela 12- Proteína <i>hub</i> da rede do interactoma oral de <i>Streptococcus mutans</i>	52
Tabela 13- Proteína de <i>Streptococcus mutans</i> com maior valor de <i>degree</i> na rede do interactoma oral de <i>Streptococcus mutans</i>	53

Agradecimentos

À Professora Doutora Maria José Correia,

Pela disponibilidade total, pelos conhecimentos transmitidos, pela motivação e incentivo durante todo o percurso académico. Por toda a paciência durante a elaboração do presente trabalho.

Ao Professor Doutor Joel Arrais,

Por toda a disponibilidade mostrada desde o primeiro momento e pelos conhecimentos transmitidos.

À Professora Doutora Marlene Barros e ao Professor Nuno Rosa,

Pelo apoio, simpatia e disponibilidade prestado na elaboração deste trabalho bem como em todo o percurso académico.

Aos meus pais,

Por todo o apoio e paciência pois sem eles nada teria sido possível. A eles devo tudo aquilo que hoje sou. Obrigado por Tudo!

À Lili,

Pela amiga, companheira e confidente nesta fase da minha vida. Por todos os momentos durante os 5 anos e pela amizade para a vida. Obrigado Binómia!

À Juliana,

Por tudo desde sempre, e nesta fase pelo apoio constante e paciência.

À Andreia,

Pela amizade e força durante esta etapa, por toda a motivação diária para alcançar a meta.

Às minha colegas de casa Joana e Rita,
Por todos os bons momentos passados juntas e também pelo apoio, força e motivação durante os 3 anos.

Ao Gonçalo,
Pela paciência, amor e por todo o apoio durante os dias todos. Pela capacidade de me pôr a sorrir nos momentos mais difíceis.

A todos os Professores, colegas e funcionários da Universidade Católica Portuguesa que me acompanharam ao longo destes 5 anos pelos conhecimentos transmitidos e por os bons momentos passados.

Lista de Abreviaturas

ADN- *Ácido desoxirribonucleico*

BioGrid- *Biological General Repository for interaction datasets*

BLAST- *Basic Local Alignment Search tool Blast*

DCT- *Discrete Casine Transform*

IntAct- *Molecular Interaction Database*

PPIs- *Protein-Protein interaction*

Uniprot- *Universal protein Resource*

Resumo

Introdução: As proteínas são moléculas essenciais ao funcionamento do organismo, apresentam inúmeras funções e estão envolvidas na regulação da maioria dos processos celulares. A interação proteína-proteína é um dos mecanismos utilizado na cavidade oral para a relação microrganismo-hospedeiro. Apesar da reconhecida importância dos microrganismos, o conhecimento específico das interações estabelecidas entre estes e o Homem é atualmente um tópico relevante de investigação, que visa compreender do ponto de vista molecular a colonização do corpo humano pelos microrganismos. Muitos resultados da proteômica têm sido publicados sobre proteínas microbianas e proteínas do hospedeiro, hoje é preciso pensar em formas de analisar e interligar os conhecimentos produzidos pelo trabalho laboratorial de forma rápida e eficaz. Com este propósito surgiram métodos computacionais para a previsão da interação proteína-proteína. A análise e interpretação crítica dos dados gerados pelas ciências Ómicas para novas propostas de interação com significado biológico só é possível com os métodos computacionais através de ferramentas bioinformáticas.

Objetivos: Analisar e discutir os resultados obtidos pelo *OralInt v2.0* à luz dos conhecimentos biológicos. Gerar informação que possibilite melhorar o desempenho do *OralInt v2.0*.

Materiais e métodos: Dos resultados obtidos do *OralInt v2.0* foram selecionadas as 200000 interações com maior nível de confiança, que foram analisadas através de diversas ferramentas informáticas.

Resultados e Discussão: O *OralInt v2.0* analisou 21086 proteínas das quais obteve um elevado número de previsão de interações. Com as 200000 interações apresentando um *score* entre 1 e 0,925 foi construído o interactoma da cavidade oral, em que constam apenas 10768 proteínas. No entanto 47% das interações são estabelecidas entre proteínas do hospedeiro e proteínas microbianas. Este resultado melhora o desempenho do *OralInt v2.0* do que o registado na versão anterior (12%). A análise topológica da rede do interactoma da cavidade oral apresenta uma distribuição de interações compatíveis com outras redes de PPIs estudadas, em que a maioria das proteínas (74%) apresentam um número de previsão de interação baixo ([1-25]). O *OralInt v2.0* prevê

algumas interações interessantes para *Porphyromonas gingivalis* e para *Streptococcus mutans* dois microrganismos importantes na cavidade oral.

Conclusão: Este trabalho mostra que a atual versão do *OralInt v2.0* já consegue uma previsão próxima do que se pensa ser o ambiente oral (há mais interações microbioma/hospedeiro previstas). No entanto são, necessários ajustes adicionais ao mesmo para que possam figurar interações entre proteínas já descritas na literatura. É ainda necessário olhar para os parâmetros analisados pelo *OralInt v2.0* para que os parâmetros biológicos das proteínas tenham um maior peso nas previsões obtidas.

Palavras-chave: Interatômica; Interação proteína-proteína; Biofilme oral; *OralInt*; *Cytoscape*

Abstract

Introduction: Proteins are essential to the organism, and many are involved in the regulation of relevant cellular processes. Protein-protein interaction is one of the mechanisms used in the oral cavity for the establishment of microorganism-host interaction. Despite the recognized importance of microorganisms, the specific knowledge of the interactions established between them and the Human is currently an important topic of research that aims to understand, from the molecular point of view, the colonization of the human body by microorganisms. Many proteomics results have been published on microbial and host proteins, now we must think of ways to analyze and interlink the knowledge produced by laboratory work. For this purpose several computational methods emerged for predicting protein-protein interactions. The analysis and critical interpretation of data generated by the Omics sciences for new proposals of interactions with biological significance is only possible with computational methods using bioinformatics tools.

Objectives: To analyze and discuss the results obtained by OralInt v2.0 in the light of biological knowledge. To generate information that will allow the improvement of the performance of OralInt v2.0.

Methods: From the results of OralInt v2.0, 200,000 interactions were selected with the highest confidence level and analyzed with various software tools.

Results and Discussion: The OralInt v2.0 analyzed 21086 proteins, and produced a large number of predicted interactions. Only the interactions with a significance level between 1 and 0,925 were analyzed. The interactome of the oral cavity, was constructed with 10,768 proteins. However, 47% of the interactions are established between the host and microbial proteins. The topological analysis of the interactome of the oral cavity presents a distribution network compatible with interactions of other PPIs networks studied, in which most proteins (74%) have a low number of prediction interaction ([1-25]). New interactions not previously described are proposed for *Porphyromonas gingivalis* and *Streptococcus mutans*.

Conclusion: Although the current version of OralInt v2.0 can make a close prediction of what is thought to be the oral environment (there are more microbiome interactions / host predicted), further adjustments are required to include interactions between

proteins described in the literature. Furthermore, it is necessary to look at the parameters analyzed by OralInt v2.0 to ensure that biological parameters of proteins are given more weight in the obtained predictions.

Keywords: Interatomic; Protein-protein interaction; Oral biofilm; OralInt; Cytoscape

Nota Prévia

Neste trabalho surgem por vezes expressões em língua inglesa, escritas em itálico, cuja tradução poderia implicar a adulteração do seu significado. Assim, optámos por incluir estes termos sem tradução pois a alteração do seu significado original poderiam levar a uma interpretação imprecisa do conceito biológico.

Ao longo do presente trabalho a palavra interatómica deve ser interpretada com o significado interação proteína-proteína.

1-INTRODUÇÃO

Durante muitos anos acreditou-se que as bactérias habitavam no planeta exclusivamente sob a forma plantónica. Esse conceito foi entendido já há 3,2 milhões de anos atrás como um modelo simplista e em vez de células que se encontram isoladas e com um desenvolvimento sésil terá sido sugerido que, as células existem em aglomerados e que esses aglomerados constituem uma comunidade, o biofilme. O biofilme define-se então como um aglomerado de células distintas aderidas a uma superfície através de uma matriz exopolisacárida que as mantém unidas e aderidas ao substrato (1).

Há quase 550 anos, *Leeuwenhoek* observou minúsculos “*animalcules*” no seu microscópio e registou a grande diversidade de tamanho e forma de células que estão presentes no biofilme oral humano (2).

Atualmente a composição do biofilme oral muito estudada e, as espécies microbianas que o constituem estão identificadas. Procura-se agora perceber quais os mecanismos moleculares específicos que os microrganismos utilizam entre eles e com o hospedeiro para estabelecer ligações que permitam a simbiose e quais são as alterações que fazem com que em determinadas condições, a simbiose se desfaça e surja a patogénese.

Existem diferentes mecanismos já descritos que dependem da interação entre os microrganismos e o hospedeiro tais como: a adesão, a emissão e receção de sinais extracelulares e o ataque das células do hospedeiro ao microrganismos e/ou mesmo a internalização dos microrganismos pelas células do hospedeiro. Todas estas interações são sobretudo mediadas por interação proteína-proteína (PPI) (3).

O estudo da interatómica (PPIs) surge como consequência do crescente aumento de informação sobre proteínas humanas e sobretudo das proteínas microbianas obtida através de métodos experimentais.

Este aumento de informação tem levado a um interesse crescente dos investigadores não apenas no conhecimento das PPI intra-específicas (interatoma de espécies ou tecidos) mas, para aumentar o conhecimento dos mecanismos de colonização e patogénese, das PPI inter-espécies (4).

Apesar deste aumento na quantidade de informação disponível, foi necessário desenvolver métodos computacionais que permitissem de forma fácil, rápida e eficaz

interligar toda a informação e o conhecimento acumulado e dele extrair significado biológico. Para isso o trabalho multidisciplinar é crucial.

Assim, com o trabalho experimental para a obtenção de informação e com o trabalho bioinformático de gestão de dados e informação é possível utilizar métodos computacionais que permitam prever novas PPIs.

Os modelos computacionais permitem ainda criar rede de PPIs que depois precisam de ser analisadas e validadas sob o ponto de vista biológico.

O *OralInt* surge de um trabalho conjunto entre os Investigadores do Departamento das Ciências da Saúde da UCP-Viseu e dos Investigadores do Instituto de Engenharia, Eletrónica e Telemática da Universidade de Aveiro e do Departamento de Informática da Universidade de Coimbra cuja primeira versão foi publicada em 2014 (5) e conta já com a versão *OralInt v2.0* que é uma versão atualizada e melhorada do *OralInt* inicial. Permite prever e criar a rede de interação proteína-proteína inter-espécie da cavidade oral. Com esta ferramenta podem vir-se a desvendar novas interações entre microrganismos bem como, perceber quais as proteínas envolvidas na interação entre o hospedeiro e os microrganismos colonizadores. Este conhecimento potencia a descoberta de novos alvos terapêuticos altamente específicos que viram a ser extremamente úteis na prática clínica.

1.1 Microbioma Humano

Quando nascemos somos apenas constituídos por células eucarióticas, mas logo desde as primeiras horas, todo o organismo (superfície da pele, cavidade oral, intestino) é colonizado por uma diversidade de bactérias, *archaea*, fungos e vírus.

Toda esta comunidade de microrganismos é designada de Microbioma Humano e, contém quase dez vezes mais células que as do nosso próprio organismo, contribuindo assim para o aumento significativo do número de genes que pode ser considerado no genoma humano numa perspectiva mais alargada (6).

Em circunstâncias normais, microrganismos do microbioma humano são comensais, ajudam a digerir os alimentos e a manter o sistema imunológico funcional, a prevenir a colonização por organismos essencialmente patogénicos, entre outras funções. No entanto, nem sempre o microbioma se encontra em equilíbrio e há muito que se sabe que os desequilíbrios no microbioma humano estão fortemente relacionados com a doença, desde doenças inflamatórias do intestino, diabetes, a resistências aos antibióticos.

Desta forma, o microbioma humano é um biomarcador para a deteção precoce de doenças bem como o alvo de intervenção terapêutica (6).

1.2 Cavidade Oral

A cavidade oral possui um conjunto de superfícies com características diferentes. Não é surpreendente que a flora oral seja uma mistura complexa de espécies microbianas que inclui bactérias, *mycoplasma*, fungos e protozoa dada a diversidade de microhabitats que se encontram neste sistema (7).

A cavidade oral apresenta duas superfícies principais que as bactérias podem colonizar: as superfícies dos dentes (tecido duros) e os tecidos da mucosa oral. São superfícies com características diferentes mas com condições física e químicas favoráveis para a adesão microbiana (8).

Assim os dentes, o sulco gengival, a língua, a mucosa jugal, o palato duro e mole são locais com colonização microbiana, com características diferentes e que levam a colonização por microrganismos também eles diferentes. Podem ainda, em casos de

pacientes desdentados, as próteses dentárias constituir um habitat para os microrganismos (9).

1.3 Ecossistema Oral

O tipo de espécies, a abundância e a atividade dos microrganismos presentes na cavidade oral vão variar em função dos parâmetros ambientais (físico, químico e biológico) bem como dos níveis de nutrientes do hospedeiro.

Assim, as comunidades microbianas orais podem apresentar grandes e rápidas mudanças na sua composição bem como na sua distribuição espacial devido à dinâmica do hospedeiro.

Estas alterações complexas e dinâmicas são consequência de vários fatores, incluindo a frequência de ingestão de comida por parte do hospedeiro, a resposta rápida à variação do pH, a interação bactéria-bactéria num período de maior tempo, as mutações genéticas e a transferência horizontal de genes entre bactérias que conferem novas propriedades aos microrganismos.

A comunidade microbiana presente no biofilme depende da combinação de fatores do hospedeiro e fatores ambientais (10).

1.3.1 Microbioma Oral

Os membros da comunidade microbiana oral desempenham um papel fundamental na manutenção da saúde oral e são os supostos agentes responsáveis pelo aparecimento e progressão das doenças orais. Estudos estimam que mais de 700 espécies de microrganismos habitem na cavidade oral constituído o microbioma oral (11).

1.4 Biofilme Oral

O biofilme oral possui uma estrutura tridimensional com uma comunidade de multi-espécies, dinâmica, com redes metabólicas e uma multiplicidade de funções interligadas (12).

O processo de formação do biofilme é altamente ordenado e coordenado, envolvendo um processo sequencial (10).

Na fase inicial forma-se a película aderida, que é colonizada por colonizadores primários e posteriormente por os colonizadores secundários.

Atualmente sabe-se que a película aderida é composta por moléculas derivadas do hospedeiro como glicoproteínas e proteínas salivares, restos de alimentos que apresentam recetores para os colonizadores primários sendo eles, *Streptococci spp*, *Actinomyces spp*, *Capnocytophaga spp*, *Eikenella spp*, *Haemophilus spp* e *Veillonella spp*.

Vulgarmente os primeiros a colonizar a cavidade oral são os *Streptococci*, especialmente *Streptococcus salivarius*. A atividade metabólica destas espécies pioneiras altera o ambiente e cria condições para a colonização da cavidade oral por parte dos outros microrganismos.

Assim que estes colonizadores formam a película aderida eles começam a excretar os seus produtos extracelulares que vão fornecer locais de ligação a outras bactérias. Subsequentemente, ligam-se bactérias como o *Fusobacterium nucleatum*, *Treponema*, *Porphyromonas gingivalis* e *Aggregatibacter actinomycetemcomitans*, sendo estes colonizadores secundários.

Com o amadurecimento do biofilme a sua composição vai-se alterando. Diminui a quantidade de *Streptococci* e *Neisseria* aumentando a quantidade de *Fusobacterium*, *Actinomyces*, *Corynebacterium*, *Veillonella*.

Para que o resto das etapas de formação do biofilme ocorra é necessário que exista comunicação entre as diferentes espécies microbianas. Importa então perceber quais os mecanismos que as bactérias utilizam para comunicarem e formar o biofilme.

Os mecanismos envolvidos na formação do biofilme oral são: adesão entre as bactérias, isto é, a presença de recetores nas membranas que levam à ligação de outras bactérias, esta adesão vai permitir à alteração do fenótipo bacteriano e das moléculas proteicas presentes à superfície; comunicação metabólica que resulta do metabolismo das bactérias que alteram as condições ambientais do biofilme e desta forma vão dar informações a outras bactérias que alteram a expressão genética e fenotípica. Por exemplo, *Streptococcus mutans* produz uma elevada quantidade de ácidos que inibe o crescimento do *Streptococcus sanguis*. Por sua vez a *Porphyromonas gingivalis* e a *Treponema denticola* apresentam um crescimento sinérgico com a produção de ácido butírico pela *Porphyromonas gingivalis* e o ácido succínico da *Treponema denticola* (13). Estes exemplos estão fortemente associados às duas grandes patologias da cavidade oral: a cárie dentária e a periodontite.

1.5 Interatômica

A palavra interatômica é utilizada na literatura com vários significados. Inicialmente definia as interações moleculares entre proteínas presentes no interior das células, atualmente a palavra interatômica é utilizada para descrever a interação proteína-proteína (PPI) no sentido mais global e inclui interações entre proteínas de células ou mesmo organismos diferentes (14, 15).

1.5.1 Interação proteína-proteína (PPIs)

As proteínas exercem a sua função através das interações entre elas, assim é importante conhecer a natureza específica destas interações.

A PPIs é um processo em que 2 proteínas se ligam fisicamente a fim de realizar uma função biológica envolvida em vários processos celulares descritos de seguida. Existem diferentes classificações para a interação entre proteínas: dependendo da força de interação, da especificidade, da localização da interação e por último de acordo com as subunidades que interagem (5).

Para que a interação se estabeleça entre as proteínas, elas devem possuir características estruturais e funcionais compatíveis, isto é, devem ser complementares (16).

As PPIs estão envolvidas em todos os processos celulares, tais como a transdução de sinal, o transporte intracelular, a replicação de ADN, a transcrição, a tradução, o *splicing*, a secreção, o controlo do ciclo celular, o metabolismo intermediário (5).

Alguns dos processos celulares mediados por PPI são processos que envolvem mais que uma célula podendo, portanto, envolver células de organismos diferentes, assim as PPIs podem ser intra-específicas ou inter-específicas.

Para estudar as PPIs a fim de se determinar o interatoma humano vários métodos experimentais têm sido utilizados como: *two-hybrid screening*, espectrometria de massa de purificação por afinidade, *microarrays* de proteína, cristalografia de raio-X e espectroscopia de ressonância magnética nuclear, transferência de energia de ressonância de fluorescência, *surface plasmon resonance*, microscopia de força atômica, microscopia eletrónica (17).

No entanto e apesar de os métodos experimentais serem o único método viável para a validação de PPIs, estes métodos apresentam grandes desvantagens que tornam o método experimental não aplicável a grande escala. As desvantagens são o tempo de execução, o custo e no caso das abordagens de alto rendimento a baixa especificidade conduzindo a um número elevado de falsos positivos.

Devido ao aumento de informações da genómica e da proteómica existe atualmente uma elevada quantidade de informação biológica que precisa de ser interligada.

De acordo com uma versão do Uniprot cerca de 135 mil proteínas humanas estão identificadas, se este número é já um problema, ele ganha outra dimensão quando se considera a interação entre proteínas humanas e microbiológicas. Estes números fazem então perceber que de facto é necessário recorrer a métodos computacionais para a determinação das PPIs.

No caso da cavidade oral, existem evidência da presença de cerca de 2300 espécies (5), assim o número de interações a ser testado neste sistema assume proporções intratáveis a não ser com grande capacidade de computação.

1.5.2 Interatoma microbioma- hospedeiro

A interação microbioma-hospedeiro é inter-espécie e pode apresentar diferentes formas, o seu conhecimento é útil para esclarecer os mecanismos utilizados na colonização e no estabelecimento da infeção (18).

A adesão ao hospedeiro é o primeiro passo para a colonização e infeção microbianas. É através do mecanismo de adesão que o microrganismo consegue chegar e colonizar o ambiente criado pelo hospedeiro. As bactérias gram-positivas e gram-negativas utilizam uma multiplicidade de proteínas que estão presentes na sua parede celular para colonizarem o hospedeiro (19).

A caracterização exata das interações estabelecidas entre o hospedeiro e o microrganismo apresenta uma elevada relevância para a definição de como a interação tem sucesso e para a compreensão do processo de estabelecimento da patologia. Um problema neste aspeto é a pouca informação existente sobre as proteínas microbianas sendo atualmente o desafio a caracterização da proteómica dos agentes microbianos (20).

A interação microrganismo-hospedeiro ocorre quando o microrganismo consegue colonizar com o hospedeiro e esta colonização passa despercebida ao sistema imunológico do hospedeiro (18).

O conhecimento do mecanismo de interação e a identificação dos locais de ligação são informações muito relevantes que permitem o desenvolvimento de terapias antimicrobianas, uma vez que se sabe que as doenças infecciosas resultam em milhões de mortes por ano (21).

O desenvolvimento dos métodos computacionais para a previsão das PPIs permitiu identificar as interações microrganismo-hospedeiro e têm levado ao avanço da descoberta dos potenciais alvos terapêuticos (22).

Vários autores realizaram estudos e desenvolveram teorias para justificar as interações patogénico-hospedeiro e em consequência a criação de fármacos altamente específicos (22).

1.6 Ferramentas bioinformáticas para o estudo do interatoma da cavidade oral

As ferramentas bioinformáticas surgem com a necessidade de filtrar e analisar toda a informação que surge das diferentes ciências Ómicas. Existem os bancos de dados primários como por exemplo *GenBank*, *Protein Identification Resource*, *EMBL Nucleotide Sequence Database*, *Gene Ontology* que são bancos onde os dados são obtidos diretamente dos métodos experimentais (23).

Surgem depois os bancos de dados secundários que são construídos a partir dos primários como exemplo *NCBI Protein*, *UniGene*, *BioGrid* (24).

O aumento da informação depositada nos bancos de dados fez com que se procure uma alternativa à interpretação e a interligação de toda a informação. Desta forma os métodos computacionais vieram trazer a ajuda que tanto se procurava na análise e interpretação dos dados obtidos.

Alguns dos métodos computacionais têm sido propostos para prever as PPIs inter-espécies. Estes métodos podem ser classificados de acordo com a informação que analisam. Uma abordagem comum é extrair a informação a partir da literatura biomédica, além deste método podem ser utilizados os dados depositados nas bases de dados de genómica, a estrutura da proteína e informação sobre o domínio e sequência das proteínas (5).

Embora a quantidade de informação da sequência genómica continue a aumentar exponencialmente, bem como a anotação da sequência das proteínas ainda temos uma limitação no acesso à informação. O que torna uma desvantagem a utilização dos métodos computacionais. Apesar desta desvantagem os métodos computacionais têm atraído a atenção dos biólogos, devido à capacidade de prever PPIs e validar os resultados obtidos por métodos experimentais por um lado e por outro por prever interações que possa ser testadas experimentalmente mas de forma mais dirigida (25).

A reconstrução rápida e precisa das redes de PPIs microrganismo-hospedeiro em todo o proteoma é essencial para revelar as interações que são estabelecidas entre o hospedeiro e o patogénico (4), e depende de ferramentas de visualização capazes de lidar com a quantidade massiva de dados.

1.6.1 *OralOme, Oralcard e OralInt*

O *OralOme* é uma ferramenta que reúne toda a informação obtida dos estudos da proteómica de proteínas humanas e proteínas microbianas presentes na saliva. Surge na necessidade de complicar a informação e fornecer uma ferramenta para os investigadores e clínicos da biologia oral. Nesta ferramenta constam marcadores moleculares específicos de determinada patologia, desenvolvimento de testes de diagnóstico e prognóstico contribuindo para a descoberta de novos agentes terapêuticos (26).

Com a informação presente no *OralOme*, surge o *OralCard*. É uma ferramenta bioinformática que resulta do trabalho conjunto entre investigadores do Departamento de Ciências da Saúde da Universidade Católica Portuguesa e do grupo Bioinformático de Aveiro que pretende com esta ferramenta informática que possa ser visto de uma forma rápida e eficaz as proteínas e os microrganismos associados às patologias orais (27).

Pode aceder-se a esta ferramenta através <http://bioinformatics.ua.pt/OralCard/> e a pesquisa pode ser feita a partir do nome da proteína oral (código *Uniprot*), através do nome da patologia (termo *MeSH* da doença) e através do nome do organismo.

O *OralInt* é um modelo computacional desenvolvido por investigadores do Instituto de Engenharia, Eletrónica e Telemática da Universidade de Aveiro e do Departamento de Informática da Universidade de Coimbra juntamente com investigadores do Departamento de Ciências da Saúde da Universidade Católica

Portuguesa. O *OralInt v2.0* utiliza as informações biológicas recolhida da base de dados do *OralCard* para prever os resultados da interação proteína-proteína inter-espécie na cavidade oral do ser humano, ambiente particularmente propenso à colonização microbiana. O *OralInt v2.0* conta já com a segunda versão. Uma das principais diferenças entre as versões é o número de proteínas utilizadas uma vez que o *OralCard* apresenta um constante atualização. Na primeira contou com um total de 3572 proteínas (2747 humanas e 825 microbianas) nesta segunda versão utilizou-se 21086 proteínas (9882 proteínas humanas e 11204 microbianas).

A determinação do interactoma oral irá permitir identificar novos biomarcadores para doenças orais e a presença de proteínas na saliva podem sugerir a saliva como um fluido de diagnóstico em alternativa ao sangue (5).

1.6.2 Cytoscape e os seus layouts como software informático para análise das redes de PPIs

Após criadas as redes de interação é necessário descodificar e interpretar as redes e para tal muitos autores utilizam o *Cytoscape*.

O *Cytoscape* é um *software* de acesso livre que permite a integração, visualização e a análise das redes biológicas (28). Assim os pares de interação determinados pelo *OralInt v2.0* por exemplo, podem ser representadas como uma rede de interações. Este *Software* para além de representar a rede permite também fazer a sua análise do ponto de vista topológico. Existem vários *plugins* para este *software* que tomam a visualização e a análise da topologia mais fácil e eficaz para utilizadores não especializados.

1.7 Análise de redes de PPIs

Após geradas as redes de interação PPI pelos métodos computacionais é essencial analisar e retirar a informação biológica contida nas mesmas.

Importa perceber que no contexto específico das redes de PPIs é preciso entender a natureza dos complexos multiproteínas, isto é, a interação entre duas proteínas pode ter consequência numa série de eventos celulares e não apenas alterar as funções das duas proteínas envolvidas na interação. Isto porque a maioria das proteínas estão envolvidos em processos dependentes de uma sequência de processos (29).

O estudo das redes de PPI exige não só o conhecimento da estrutura da proteína mas também a informação funcional é essencial para a compreensão completa da rede gerada.

É necessária a identificação das proteínas essenciais contidas na rede assim como das propriedades topológicas das mesmas de acordo com os valores de centralidade (30).

Um método para identificar proteínas essenciais a partir de uma rede, passa pela ideia de que cada proteína dentro da rede em que está inserida é vista como uma unidade, que estabelece interações com as outras proteínas presentes na rede, desta forma é possível calcular os valores topológicos da proteína dentro da rede. Uma das ferramentas informáticas que pode analisar os valores topológicos referentes a cada proteína na rede em que está inserida é o *Cytoscape* (30).

Desta forma recentemente os investigadores vieram a provar a correlação positiva entre as proteínas essenciais e os valores topológicos de centralidade (31).

As métricas existentes para avaliar a topologia da rede são o *degree*, *closeness centrality*, *hub*. Todos estes parâmetros atribuem um valor numérico a cada proteína contida na rede (32, 33).

Vários estudos recentes estabelecem uma relação entre o valor de *degree* e a importância das proteínas, isto é, proteínas com um valor de *degree* maior são proteínas de importância acrescida dentro da rede em estudo (32).

O valor de *degree* pode ser definido como o número de ligações (*edges*) que cada nó (*nodes*) estabelece (34). Num contexto de uma rede de interação de proteínas o valor de *degree* define-se como o número de interações que cada proteína estabelece dentro da rede em estudo.

As proteínas com maior valor de *degree* são consideradas proteínas *hubs*, isto é, são as proteínas estabelecem um número de interações muito superiores face às restantes proteínas da rede. (34).

2-OBJETIVOS

- Analisar e discutir os resultados obtidos do *OralInt v2.0* da previsão de interação proteína-proteína na cavidade oral após as alterações executadas no *OralInt v2.0*;
- Utilizar a ferramenta *Cytoscape* de forma adequada e eficaz para a análise da rede do interactoma da cavidade oral;
- Descrever as redes de interações estabelecidas entre as proteínas pertencentes às diferentes espécies focando nas interações entre as proteínas humanas (hospedeiro) e as proteínas microbianas (patogénico e simbiote);
- Verificar se há novas propostas de interação obtidas pelo *OralInt v2.0* não descritas na literatura, e que tenham sentido biológico;
- Permitir com as interações geradas pelo *OralInt v2.0* e depois de analisadas e interpretadas esclarecer a relação entre as espécies presentes no microbioma oral;
- Verificar se as interações propostas permitem sugerir novos mecanismos fisiopatológicos;
- Aplicar as informações obtidas à proposta de novas estratégias de terapêutica e/ou diagnóstico;
- Gerar um conjunto de sugestões que possibilitem melhorar o desempenho do *OralInt v2.0*.

3-MATERIAS E MÉTODOS

3.1 Proteoma Oral

Como ponto de partida para o nosso estudo, foram utilizadas um total de 21086 proteínas, 11204 proteínas microbianas e 9882 proteínas humanas, depositadas no *OralCard* (Dezembro 2013).

As proteínas foram identificadas através da análise de proteómica da saliva, utilizando métodos como a 2D electroforese/espectrometria de massa ou a 2D cromatografia líquida/espectrometria de massa (5).

Para a criação do *OralInt v2.0* a recolha dos dados biológicos foi realizada a partir das bases de dados do Uniprot e da BioGrid.

O Uniprot foi utilizado com o propósito de extrair informação funcional de cada proteína, como o seu nome, descrição geral, sequência de aminoácidos, conjunto de anotações ontológicas, entre outras.

Por sua vez a BioGrid, permitiu ter acesso a um reportório de informação vastíssimo sobre interações das proteínas e a sua ligação a patologias. Este banco de dados permitiu o acesso a interação de proteínas de múltiplos microrganismos. Com esta estratégia foi construída uma base de dados teste com os pares de interações positivas. Os dados testes positivos utilizados pelo *OralInt v2.0* foram as interações já previstas e depositadas no BioGrid apenas por métodos experimentais.

Um dos principais problemas encontrado foi a criação do conjunto de pares de interações negativos, ou seja, interações que não são possíveis de ocorrer entre as proteínas. Este tipo de interações não é frequentemente publicado pelos investigadores o que torna numa dificuldade acrescida. Para ultrapassar este obstáculo, o grupo de interações negativas foi criado tendo por base as propriedades físico-químicas das proteínas, isto é, foram criados pares de interações que devido à incompatibilidade físico-química existente entre as proteínas a sua interação será impossível.

Apesar de uma descrição extensiva dos detalhes do funcionamento do *OralInt v2.0* de previsão poder ser encontrado em (35) o método computacional utilizado consiste na utilização e no treino de uma máquina de vetores de suporte. A transformação das sequências num subespaço de *features* foi implementada recorrendo a uma técnica designada de DCT (*Discrete Cosine Transform*).

Para a utilização deste método é necessário a extração das características da proteína a partir da sua sequência de aminoácidos. Esta sequência é depois convertida

em categorias físico-químicas de acordo com uma tabela proposta por *Shen et al.* sendo posteriormente aplicada a DCT. Ele vai olhar para o conjunto de dados da proteína a treinar e classificar olhando apenas para a sequência da proteína. Por outras palavras usa apenas a sequência da proteína, apesar da existência de outra informação que neste caso é a informação biológica que permite melhorar o seu desempenho (35).

Dos resultados obtidos do *OralInt v2.0* selecionaram-se as 200000 interações com nível de confiança mais elevado.

3.2 Análise das redes geradas

O programa informático selecionado para visualizar e validar as interações entre proteínas foi o *Cytoscape* (<http://www.cytoscape.org/>). O *Cytoscape* disponibiliza diversos *layouts* para visualização das redes.

Foi importado um documento em formato Excel onde consta o par de interação e o *score* atribuído a cada par de interação. Foi definido que a cada linha da folha de Excel correspondia uma interação entre duas proteínas, uma seria a fonte de interação (Proteína 1) e outra, o alvo de interação (Proteína 2). O tipo de interação foi definido como o *score* que cada par de interação possui (*Score*). O valor de *score* foi um valor que o *OralInt v2.0* forneceu como sendo o valor de *score* para cada par de interação. O *score* de cada par de interação corresponde à probabilidade de ocorrência da interação atribuída pelo *OralInt v2.0*. Não representa a probabilidade de interação do par na realidade biológica mas sim a probabilidade de ocorrência atribuída pelo *OralInt v2.0*.

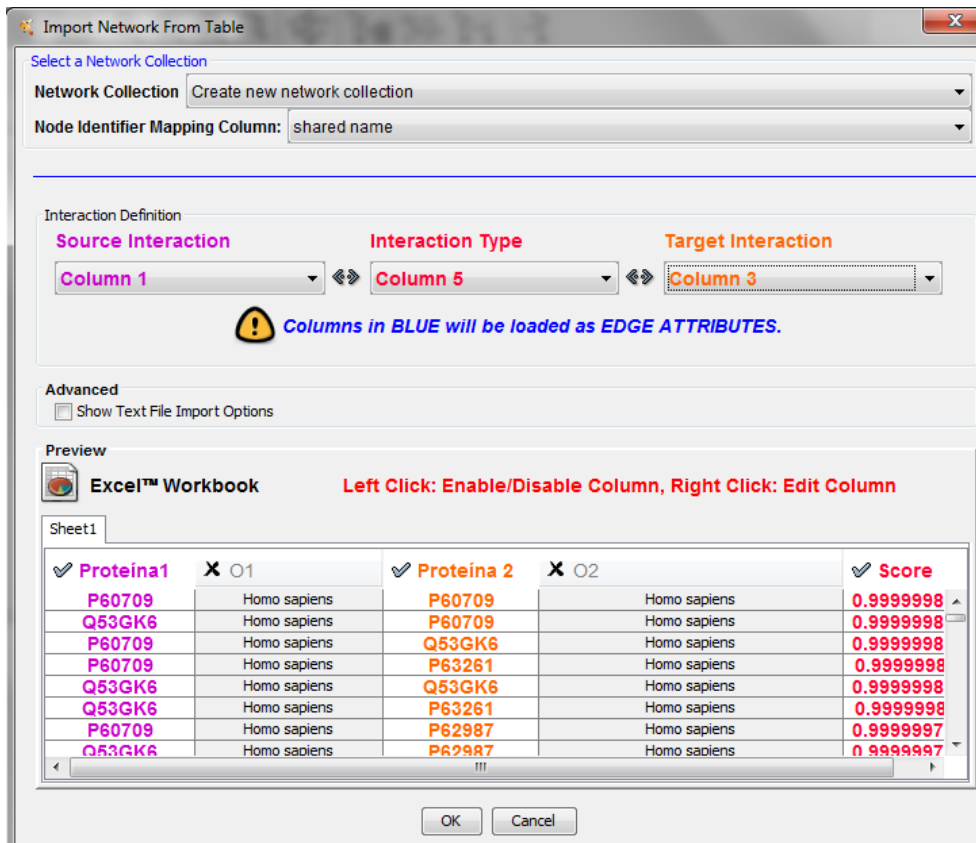


Figura 1-Interações previstas pelo *OralInt v2.0* importados para o *Cytoscape*, com a definição da fonte e alvo de interação, assim como a classificação do tipo de interação.

Foi por nós criada uma tabela de atributos para cada proteína. Para tal da lista total de proteínas foram removidos os duplicados, ficando apenas as proteínas únicas.

A cada código Uniprot de proteína foi associado o nome do organismo que a produzia (*taxon ID*) o nome da proteína, se a proteína tinha sido proposta como biomarcador e as doenças para as quais a proteína já foi identificada. Esta informação foi importada para o *Cytoscape* em formato txt.

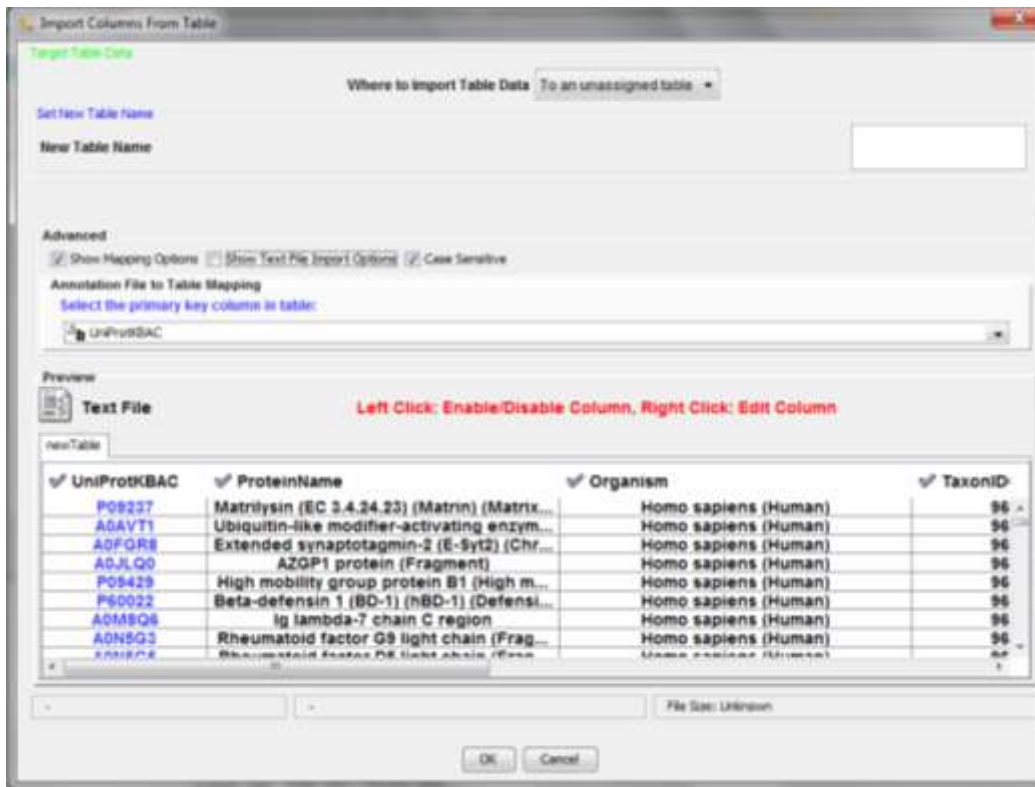


Figura 2- Atributos de cada proteína. Importação para o *Cytoscape* referentes à informação pertencente a cada proteína presente na rede.

3.3. Utilização do *Cytoscape* para análise da rede

A utilização de ferramentas disponibilizadas pelo *Cytoscape* permite uma análise topológica da rede, gerada pelas 200000 interações selecionadas para a realização deste trabalho.

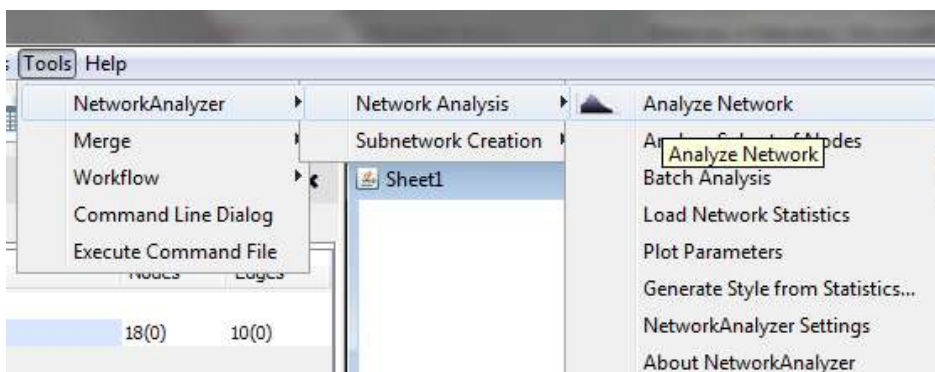


Figura 3- Ferramentas que o *Cytoscape* disponibiliza para a análise topológica da rede.

3.4 Criação de sub-redes para interpretação biológica da rede do interactoma oral

Dada a complexidade da rede do interactoma oral, foi necessário pensar uma forma de ser mais fácil extrair significado biológico a partir do total da rede do interactoma oral. A cárie dentária e a doença periodontal são duas das patologias com maior impacto na prática clínica e sabe-se que o principal agente responsável da cárie dentária é o *Streptococcus mutans* e a espécie responsável pelo estabelecimento da periodontite é a *Porphyromonas gingivalis*. Tendo em conta estes dados decidiu-se construir o interactoma oral dos dois patogénicos.

Para a construção das redes, consultou-se o documento em Excel onde constavam as 200000 interações que compõem o interactoma oral e selecionaram-se todas as interações em que as proteínas pertencentes às duas espécies estavam envolvidas, desta forma obteve-se o interactoma oral de *Porphyromonas gingivalis* e a rede do interactoma oral de *Streptococcus mutans*.

Existem já descritos na literatura fatores de virulência associados a cada uma das espécies como tal foram procurar-se as interações previstas pelo *OralInt v2.0* com as proteínas: *gingipains* e *fimA* para a *Porphyromonas gingivalis*, e a *glucosyltransferase* para o *Streptococcus mutans*.

Para localizar as proteínas (*gingipains* e *FimA*) na rede, consultou-se o Uniprot. Através do nome genérico das proteínas e pesquisou-se todas as entradas referentes a ambas as proteínas produzidas pela espécie *Porphyromonas gingivalis* (Tabela 1 e 2).

Tabela 1- Códigos Uniprot correspondentes a proteínas *gingipains* produzidas pela espécie *Porphyromonas gingivalis*

Códigos Uniprot			
<u>P95493</u>	<u>P28784</u>	<u>G1UBT8</u>	<u>Q9R512</u>
<u>B2RLK2</u>	<u>Q51817</u>	<u>F5XDI5</u>	<u>Q9R513</u>
<u>A4PCG0</u>	<u>P72197</u>	<u>B2RMI8</u>	<u>Q6Q4T4</u>
<u>Q7WUA4X</u>	<u>P72194</u>	<u>B2RIS2</u>	<u>Q51816</u>
<u>Q9ZNF9</u>	<u>Q7WUA4</u>	<u>T2N7N5</u>	<u>Q51844</u>
<u>Q9ZNF7</u>	<u>B2RHG9</u>	<u>T2N915</u>	<u>Q7WUA4</u>

Tabela 2- Códigos Uniprot correspondentes a proteínas *FimA* produzidas pela espécie *Porphyromonas gingivalis*

Códigos Uniprot		
<u>P0C940</u>	<u>P59914</u>	<u>E9LVG7</u>
<u>B2RH54</u>	<u>E9LVG7</u>	<u>T2NAH0</u>
<u>Q9S0W8</u>	<u>E9LVG6</u>	<u>T2N9X4</u>
<u>Q51827</u>	<u>R4X5V0</u>	<u>T2NA31</u>
<u>Q51826</u>	<u>Q93R80</u>	<u>R4WWB8</u>
<u>Q51822</u>	<u>Q51825</u>	<u>O32387</u>
<u>G1UBU7</u>	<u>T2N8A9</u>	<u>F4ZSI2</u>
<u>R4X5V0</u>	<u>E9LVG6</u>	

Após a recolha dos códigos Uniprot das proteínas foi procurar-se na base do *OralOme* todos os códigos apresentados das tabelas 1 e 2, para identificar os códigos que o *OralInt* v2.0 utilizou para a previsão das interações do interactoma da cavidade oral.

Para localizar as *glucosyltransferase* na rede, consultou-se a base do *OralOme* e pesquisou-se todas as entradas referentes à proteína produzidas pela espécie *Streptococcus mutans* (Tabela 3).

Tabela 3- Códigos Uniprot da proteína *glucosyltransferase* produzidas pela espécie *Streptococcus mutans*.

Códigos Uniprot
P08987
P13470
P49331

Após a recolha dos códigos da proteína *glucosyltransferase* foi procurar-se na rede do interactoma oral de *Streptococcus mutans* as previsões de interações para as três proteínas.

3.5 Ferramentas Bioinformáticas para recolha de informação funcional e biológica das proteínas

Existe um número significativo de ferramentas bioinformáticas com informação referente às proteínas descritas na literatura. A consulta das ferramentas permite hoje, aceder à informação de forma mais rápida e eficaz por parte dos utilizadores, uma vez que as ferramentas apresentam a informação sob uma forma esquemática.

A utilização de ferramentas bioinformáticas para consulta permite aos utilizadores a recolha de informação sobre as proteínas como: o gene a que pertencem, o organismo, as funções biológicas e moleculares que já lhe foram atribuídas bem como também as interações que as proteínas estabelecem com outras proteínas sejam elas pertencentes à mesma espécie ou a espécies diferentes. Das várias ferramentas disponíveis para a recolha de informação das proteínas neste trabalho utilizou-se o Uniprot, o BioGrid e o PubMed.

O Uniprot é um site de livre acesso, que permite obter a sequência das proteínas bem como a informações funcionais, contém uma grande quantidade de informação biológica das proteínas retirada da literatura e organizada de forma de fácil consulta, permitindo a exportação dos dados das diferentes proteínas.



Figura 4 -Página inicial do Uniprot obtida através do site <http://www.uniprot.org/> para recolha de informação das proteínas. Exemplo: pesquisa da proteína Q99616 onde consta a informação referente à proteína em questão.

A BioGrid (*Biological General Repository for Interaction Datasets*) permite de forma livre ter acesso às interações de proteínas humanas bem como de outros microrganismos já descritas na literatura.



Figura 5- Página inicial do BioGrid obtida através do site <http://thebiogrid.org/> para consulta de interações de proteínas já descritas na literatura.

O PubMed permitiu ter acesso a artigos científicos de onde é possível retirar informação referente aos vários tópicos citados ao longo do presente trabalho.

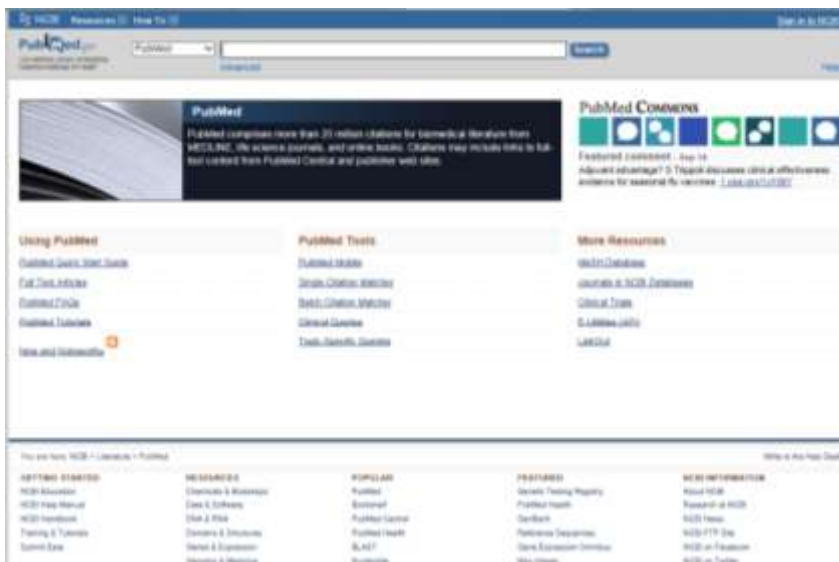


Figura 6 -Página inicial do Pubmed obtido do site <http://www.ncbi.nlm.nih.gov/pubmed/>.

4- RESULTADOS E DISCUSSÃO

4.1 Interactoma da cavidade oral

4.1.1 Proteínas usadas pelo *OralInt v2.0*

Como ponto de partida para o estudo foram utilizadas 11204 proteínas de origem microbiana e 9882 proteínas de origem humana que perfaz um total de 21086 proteínas depositadas no *OralCard* (Dezembro de 2013).

Das 11204 proteínas de origem microbiana os géneros com maior número de proteínas são apresentados na figura 7.

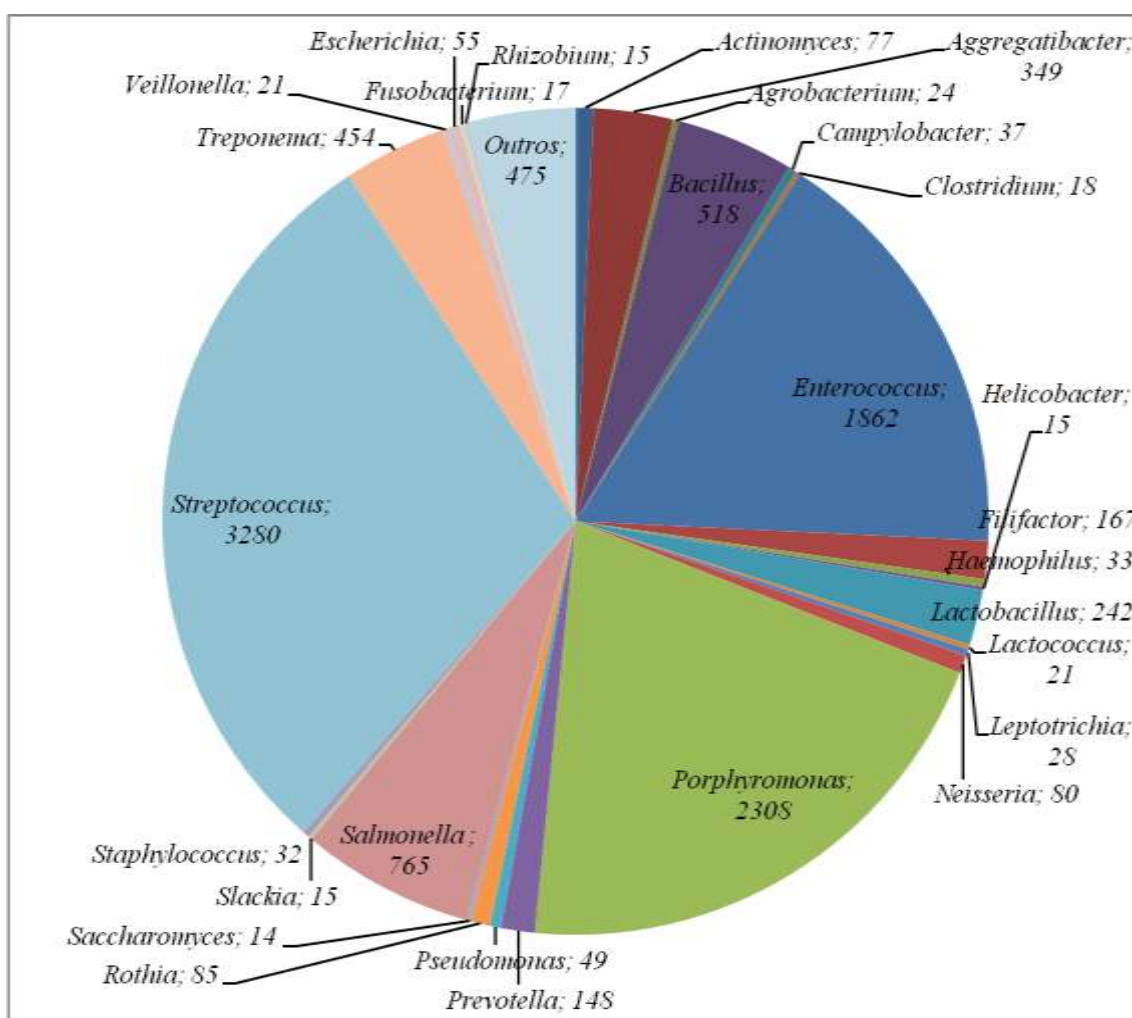


Figura 7- Número de proteínas microbianas distribuídas por género microbiano presentes na cavidade oral que o *OralInt v2.0* analisou.

Da totalidade das proteínas microbianas analisadas pelo *OralInt v2.0*, estão representados 140 géneros microbianos que colonizam a cavidade oral.

Na figura 7 a fatia do gráfico que se encontra legendada com “Outros” representa 475 proteínas de géneros microbianos diferentes onde o número total de proteínas pertencentes a cada género é menor ou igual a 10 proteínas.

4.1.2 Proteínas presentes na rede do interactoma oral em estudo

A partir das 21086 proteínas o *OralInt v2.0* previu um elevado número de interações, das quais foram selecionadas as 200000 interações, que apresentam valores de *score* compreendido entre 1 e 0,925. A média do valor de *score* das 200000 interações é de 0,954.

Nas 200000 interações foram identificadas apenas 3162 proteínas de origem humana e 7606 proteínas de origem microbiana que perfaz um total de 10768 proteínas.

Este resultado indica que no conjunto de interações selecionadas, para 32% (6720) de proteínas humana e 68% (3598) de proteínas microbianas o *OralInt v2.0* não previu interações de alta confiança (*score* > 0,925).

As 7606 proteínas de origem microbiana que estabelecem interações de alta confiança representam 134 géneros microbianos. Os géneros com maior número de proteínas estão identificados na figura 8.

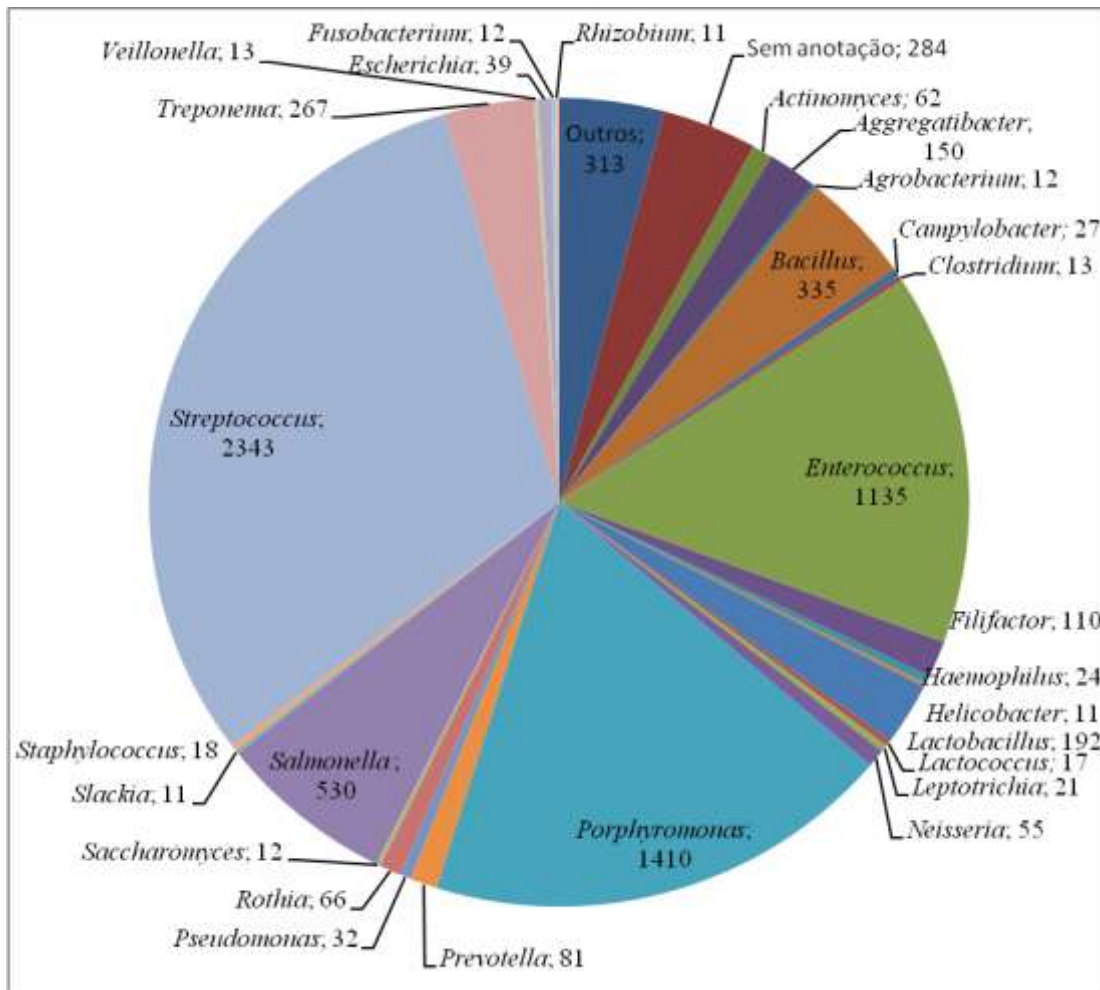


Figura 8 -Número de proteínas de origem microbiana distribuídas por gênero microbiano presentes nas 200000 interações previstas pelo *OralInt v2.0* com score superior a 0,925.

Como é possível observar na figura 8 as 7606 proteínas de origem microbiana que estabelecem interações de alta confiança, pertencem predominantemente aos gêneros *Streptococcus* (2343 proteínas), *Porphyromonas* (1410 proteínas) e *Enterococcus* (1135 proteínas).

A predominância destes gêneros na cavidade oral reflete os gêneros associados às três patologias orais mais comumente tratadas na prática clínica. Os *Streptococcus* relacionados com a cárie dentária, as *Porphyromonas* com a periodontite e o gênero dos *Enterococcus* presentes nas infecções endodônticas.

A fatia do gráfico que se encontra legendada com “Outros” representa 313 proteínas de gêneros microbianos diferentes onde o número total de proteínas pertencentes a cada gênero é menor ou igual a 10 proteínas.

A fatia do gráfico que se encontra legendada “sem anotação” representa 204 proteínas relativamente às quais não se encontrou qualquer tipo de anotação no Uniprot à data de realização do presente trabalho.

Estas proteínas foram eliminadas da base de dados, no entanto não foi possível saber-se o porquê de serem eliminadas, uma vez que o Uniprot apenas dá a informação que as entradas das proteínas foram apagadas.

A eliminação de entradas de proteínas pelos revisores do Uniprot pode ser devido a uma de várias causas, tais como, uma proteína proposta e atribuída uma entrada mas que posteriormente com revisão manual se tenha verificado que representava apenas um fragmento de uma proteína já existente. Nestes casos a entrada do Uniprot é eliminada e é atribuída apenas a entrada da proteína total. Pode também acontecer, que o que se julgava ser uma proteína única ser em trabalhos posteriores identificada como duas proteínas distintas, assim em vez de uma única entrada são atribuídas duas entradas no Uniprot que correspondem cada uma a sua proteína (36) (37).

Da análise da figura 7 e 8 a principal diferença a registar é o número total de proteínas, o que significa que para muitas proteínas (51%) o *OralInt v2.0* não prevê interações com um *score* acima de 0,925. No entanto os géneros com maior número de proteínas analisadas pelo *OralInt v2.0* (*Streptococcus*, *Porphyromonas*, *Enterococcus*) são os mesmos que estabelecem interações com *score* superior a 0,925.

A ausência de 51% de proteínas nas 200000 interações selecionadas não significa que elas não estabeleçam interações mas apenas que as interações previstas não têm um *score* suficiente para serem incluídas nas 200000 com maior *score*.

4.1.3 Tipo de interações previstas pelo *OralInt v2.0*

Das 200000 interações previstas pelo *OralInt v2.0* com *score* superior a 0,925, 94503 (47%) correspondem a interações entre proteínas humanas-microbianas, 66307 (33%) interações entre proteínas microbianas e 32143 (16%) interações entre proteínas humanas como se apresenta na figura 9.

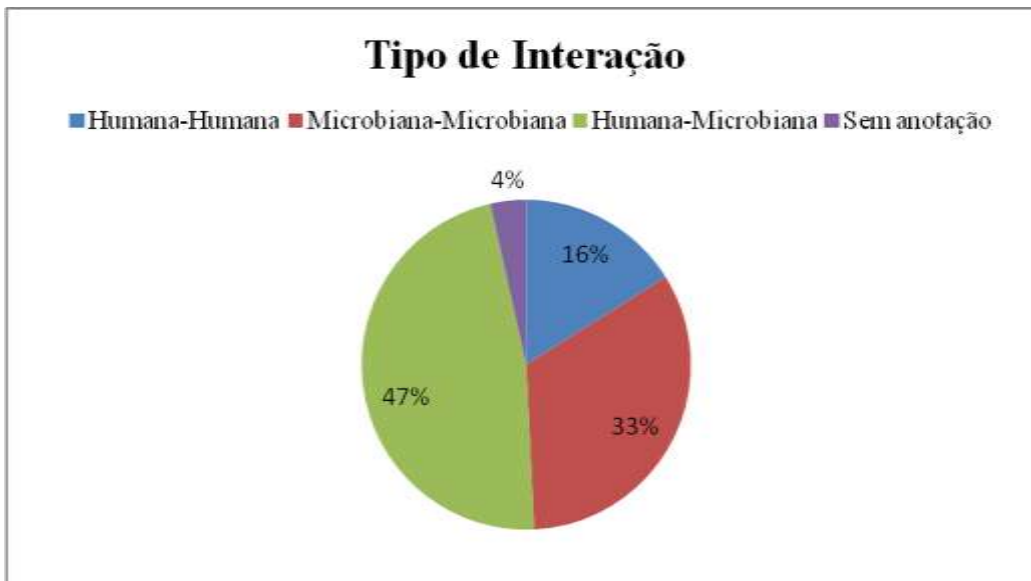


Figura 9- Distribuição da porcentagem dos tipos de interações estabelecidas entre as proteínas presentes nas 200000 interações selecionadas.

Importa salientar da figura 9 que 4% das interações estão legendadas como “sem anotação”, este resultado reflete as interações estabelecidas entre as proteínas às quais não se encontra atribuído nenhuma anotação, uma vez que, foram eliminadas do Uniprot.

Pode verificar-se que o maior número de interações é o estabelecido entre as proteínas humanas e microbianas (47%).

O *OralInt v2.0* apresenta valores diferentes do que na versão anterior (5) no que se refere à previsão de interação entre proteínas humanas e microbianas, o que melhora o seu desempenho para o estudo do interactoma da cavidade oral pois deteta um maior número de interações entre proteínas do hospedeiro e proteínas microbianas. Na primeira versão são apresentadas um total de 12% de interações entre proteínas humanas-microbianas (38) e nesta versão existem mais que o dobro deste tipo de interações (Figura 9).

4.1.4 Análise topológica da rede do interactoma oral

A rede do interactoma oral apresenta 10768 nós (*nodes*) e 200000 interações (*edges*).

No sentido de interpretar de forma global a rede de interações obtida, fez-se uma análise topológica da rede. Para tal optou-se pela seleção da ferramenta *analyse network* do *Cytoscape* que permite a análise de redes biológicas e o respetivo cálculo de parâmetros topológicos. Esta ferramenta permite a obtenção de diferentes parâmetros como o *degree*, *clustering coefficients* entre outros.

No presente trabalho olhou-se em concreto para o valor de *degree* atribuído a cada proteína representada na rede (Figura 10).

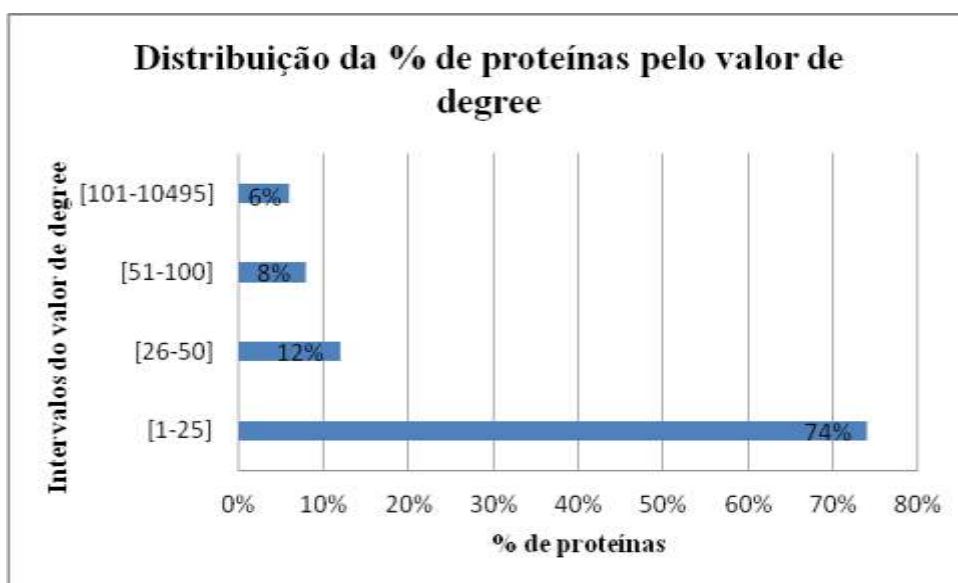


Figura 10- Representação da percentagem das proteínas agrupadas de acordo com o seu valor de *degree* na rede em estudo

O que se pode ver representado na figura 10 é que 74% das proteínas presentes na rede apresentam um valor de *degree* compreendido no intervalo [1-25], 12% apresentam um valor de *degree* compreendido [26-50], 8% das proteínas apresentam um *degree* compreendido [51-100] e 6% das proteínas apresentam um valor de *degree* [101-10495].

Este resultado é idêntico ao observado noutros estudos (34) que mostram também, que do total das proteínas presentes nas redes de PPIs, normalmente poucas são as proteínas *hub*, isto é, poucas são as proteínas que apresentam previsão de um número de interações muito superior face às restantes proteínas presentes na rede. No caso do interactoma oral apenas 6 % das proteínas apresentam previsões de interação superior a 101.

Estas são consideradas as proteínas *hub* e a Tabela 4 apresenta, as 10 proteínas *hub* da rede do interactoma oral.

Tabela 4- Proteínas *hub* da rede do interactoma oral.

Código Uniprot	Nome da proteína	Espécie	Valor de <i>degree</i>
<u>P60709</u>	<i>Actin, cytoplasmic 1</i>	<i>Homo sapiens</i>	10495
<u>Q53GK6</u>	<i>Beta actin</i>	<i>Homo sapiens</i>	10433
<u>Q9UE58</u>	<i>Delta-globin</i>	<i>Homo sapiens</i>	7395
<u>Q9BYX7</u>	<i>Actin</i>	<i>Homo sapiens</i>	5878
<u>P0A2H5</u>	<i>Protein translocase TatE</i>	<i>Salmonella typhimurium</i>	4522
<u>C5CJ16</u>	<i>RND efflux system</i>	<i>Variovorax paradoxus</i>	3754
<u>P63261</u>	<i>Actin, cytoplasmic 2</i>	<i>Homo sapiens</i>	3692
<u>P62987</u>	<i>Protein ribosomal 60S</i>	<i>Homo sapiens</i>	3158
<u>Q9S3R9</u>	<i>Membrane protein 41</i>	<i>Porphyromonas gingivalis</i>	2851
<u>B2RIO3</u>	<i>Membrane protein 41</i>	<i>Porphyromonas gingivalis</i>	2850

Da análise da tabela 4 podemos ver que as duas proteínas com maior valor de *degree* são de origem humana e apresentam um valor bastante superior aos das restantes proteínas. Estas duas proteínas (P60709 e Q53GK6) são duas actinas e na lista das 10 proteínas com mais interações previstas pelo *OralInt v2.0* encontram-se mais duas actinas (Q9BYX7 e P63261).

As actinas são um grupo de proteínas presentes no citoesqueleto de todas as células eucariotas. São proteínas essenciais para diversos processos celulares como a citocinese e a contração muscular. Desempenhando portanto um papel fundamental na organização do citoesqueleto bem como em processos celulares básicos (39). As actinas apresentam inúmeras interações descritas da base de dados da BioGrid bem como na base de dados IntAct.

No entanto após a análise da função das actinas, verificamos que são proteínas presentes em todas as células, com funções estruturais *housekeeping* e não com funções específicas dos tecidos orais. Pensando nas funções estruturais das actinas elas apresentam um papel crucial na configuração celular e estas através da ligação a outras proteínas com as integrinas permite a sua ligação a uma grande diversidade de moléculas (40).

A grande diversidade de proteínas com as quais as actinas interagem e os vários domínios de adesão que apresentam, podem justificar o elevado número de interações previstas pelo *OrallInt v2.0* para estas proteínas.

A proteína microbiana com maior valor de *degree* é uma *protein translocase* (P0A2H5) pertencente à espécie *Salmonella typhimurium*. Esta proteína transmembranar está associada ao transporte de proteínas através da membrana (41) e portanto o facto de estabelecer várias PPIs parece lógico. No entanto a *Salmonella* é um microrganismo que coloniza tipicamente o sistema gastrointestinal (42) e não a cavidade oral. Assim foram procuradas proteínas homólogas da P0A2H5 tendo-se verificado que não existia homólogos desta proteína em espécies tipicamente orais (*Streptococci*, *Porphyromonas*, *Aggregatibacter*, *Treponemas*, *Rothia*, *Fillifactor* e *Enterococcus*).

Dentro da lista das 10 proteínas *hub*, encontram-se ainda duas proteínas membranares da espécie *Porphyromonas gingivalis*. Este é o resultado mais interessante do ponto de vista do estudo da interactómica oral, dado que a espécie *Porphyromonas gingivalis* coloniza a cavidade oral e sabe-se que a sua presença está relacionada com o estabelecimento da doença periodontal (43). A proteína de *Porphyromonas gingivalis* que estabelece mais interações é a Q9S3R9 uma proteína membranares cuja função é ser um transportador de membrana com funções de ligação de polissacarídeos e de porina (41). Esta proteína tem uma sequência idêntica à da proteína B2RIQ3 havendo apenas um aminoácido diferente em 391 presentes. Isso justifica, que o número de interações previstos pelo *OrallInt v2.0* seja praticamente o mesmo. A diferença entre estas duas proteínas é o facto da primeira se encontrar revista na base de dados Uniprot enquanto que a segunda ainda não foi revista manualmente.

Da análise global da tabela 4 verifica-se que estão presentes uma maioria de proteínas humanas (6/10) e que das proteínas microbianas apenas duas pertencem a microrganismos da cavidade oral. As proteínas humanas presentes têm funções *housekeeping* nas células do hospedeiro e não funções associadas à interação com o microbioma. Quanto às proteínas microbianas presentes, independentemente da espécie bacteriana a que pertencem, estão sempre associadas a transporte membranares.

Quando se procura o significado das proteínas *hub* nas redes de interactómica é curioso perceber que no caso de uma rede PPI de leveduras se ocorrer a remoção das proteínas *hub*, o que têm sido demonstrado empiricamente é que as leveduras apresentam menor capacidade de sobreviver. Por sua vez, numa rede de PPI humanas as proteínas

hub, são aquelas que estão envolvidas em processos celulares que ocorrer em todas as células, isto é, processos estruturais da própria célula, enquanto que, as proteínas que estão envolvidas em processos mais específicos de cada tecido estabelecem um menor número de interação. Nas rede de PPI no geral as proteínas *hub*, podem ser aquelas que dadas as suas características estruturais apresentam maior capacidade de interagir com outras proteínas (44).

As proteínas *hub* da rede do interactoma oral, apresentam o mesmo comportamento do que o verificado noutros estudos (44) das proteínas *hubs* de rede de PPI humanas e também das redes de PPIs no geral. De acordo com este resultado podemos ver que, a distribuição topológica da rede do interactoma oral é idêntica a outras redes de PPIs estudadas, o que se torna um resultado positivo das previsões obtidas do *OralInt v2.0*.

Dado que o interactoma oral total revelou ser uma rede demasiado complexa, consegue-se extrair pouco significado biológico da sua análise global. Foram portanto selecionadas as proteínas envolvidas em duas patologias orais com maior impacto na prática clínica (a cárie e a doença periodontal) para compreender qual a proporção que estas interações representam na rede total e o que as interações previstas revelam dos mecanismos que ocorrem neste habitat.

4.2 Interações previstas para proteínas de microrganismos associados a patologias orais

Do total de proteínas analisadas e das interações selecionadas do *OralInt v2.0* agruparam-se as proteínas em 4 grupos. O grupo “*Homo sapiens*” onde se encontram representadas todas as proteínas de origem humana, o grupo “cárie” onde estão representadas as proteínas pertencentes às espécies listadas na tabela 5 que representam espécies microbianas associadas à cárie dentária, o grupo “perio” onde estão representadas as proteínas das espécies listadas na tabela 5 de microrganismos envolvidos na doença periodontal e no grupo “Proteínas microbianas” estão todas as restantes proteínas que não estão intimamente relacionadas com as duas grandes patologias da cavidade oral (cárie e doença periodontal). A proporção de interações estabelecida entre estes 4 grupos de proteínas é apresentada na figura 11.

Tabela 5- Espécies microbianas presentes na amostra do *OralInt v2.0* associadas às duas patologias orais com maior impacto na medicina dentária (cárie e doença periodontal).

Espécies envolvidas na cárie dentária	Espécies envolvidas na doença periodontal
<i>Actinomyces odontolyticus</i>	<i>Aggregatibacter actinomycetemcomitans</i>
<i>Actinomyces sp.</i>	<i>Campylobacter concisus</i>
<i>Anaerococcus tetradius</i>	<i>Campylobacter curvus</i>
<i>Bifidobacterium dentium</i>	<i>Eikenella corrodens</i>
<i>Lactobacillus acidophilus</i>	<i>Filifactor alocis</i>
<i>Lactobacillus fermentum</i>	<i>Porphyromonas endodontalis</i>
<i>Streptococcus mutans</i>	<i>Porphyromonas gingivalis</i>
<i>Streptococcus sobrinus</i>	<i>Porphyromonas uenonis</i>
	<i>Prevotella bivia</i>
	<i>Prevotella buccalis</i>
	<i>Prevotella intermedia</i>
	<i>Prevotella oralis</i>
	<i>Prevotella oris, Prevotella sp.</i>
	<i>Prevotella tanneriae</i>
	<i>Prevotella veroralis</i>
	<i>Rothia mucilaginosa</i>
	<i>Tannerella forsythia</i>
	<i>Treponema denticola</i>
	<i>Treponema lecithinolyticum</i>
	<i>Treponema pallidum</i>
	<i>Treponema socranskii</i>
	<i>Treponema vincentii</i>

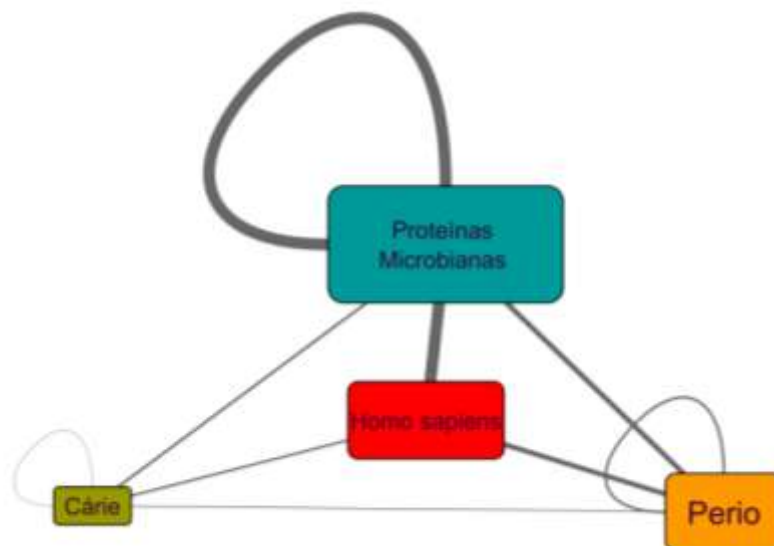


Figura 11- Representação das interações previstas pelo *OralInt v2.0* entre os 4 grupos: as proteínas humanas (vermelho), proteínas de espécies microbianas associadas à cárie (verde), proteínas de espécies microbianas associadas à doença periodontal (laranja) e as restantes proteínas microbianas (azul). O tamanho da caixa é proporcional ao número de proteínas pertencentes a cada grupo. As interações entre as proteínas dos diferentes grupos são representadas pelas linhas de forma proporcional ao número de interações previstas pelo *OralInt v2.0*.

Verifica-se que o grupo “Proteínas Microbianas” é o grupo que apresenta um maior número de proteínas e o grupo “cárie” é o grupo que apresenta um menor número de proteínas. O que significa que a maior parte das proteínas presentes neste estudo pertencem a espécies microbianas que ainda não foram associadas à cárie dentária ou à doença periodontal e que podem não estar associadas a patologias mas sim ao biofilme saudável.

Por sua vez também a espessura das linhas que representa as interações entre as proteínas pertencentes aos diferentes grupos é proporcional ao número de interações geradas pelo *OralInt v2.0*. Assim, como é possível ver na figura 11 a linha de maior espessura é a que se encontra no grupo “Proteínas microbianas” o que significa que das 200000 interações o maior número de interações ocorre entre espécies microbianas que não foram incluídas nos dois grupos pertencentes à “cárie” e “perio”. A linha mais espessa que se encontra a ligar dois grupos é a que representa as interações entre as proteínas humanas e as proteínas microbianas pertencente ao grupo “Proteínas microbianas”. A identificação e estudo detalhado deste tipo de interação é fundamental para a compreensão do biofilme oral comensal.

Não está representada a linha de interação entre as proteínas humanas uma vez que no presente trabalho pretende-se focar o estudo nas interações inter-espécie, mas tal como representado na figura 9 as interações entre as proteínas humanas representam 16% do total de interações analisadas neste trabalho.

4.2.1 Interactoma oral de *Porphyromonas gingivalis*

Dado que na secção 4.2 foi verificado que as interações estabelecidas por proteínas produzidas por periodontopatogénicos são numerosas (96980) e no sentido de verificar se algumas das interações descritas na literatura eram reconhecidas pelo *OralInt v2.0* dentro do intervalo de previsões utilizado para o estudo, selecionou-se um periodontopatogénico bem estudado, a *Porphyromonas gingivalis*. Da rede total do interactoma oral em estudo foram selecionadas apenas o grupo de interações em que estão envolvidas proteínas de *Porphyromonas gingivalis*.

A rede do interactoma oral de *Porphyromonas gingivalis* apresenta 43210 interações previstas, estando representas 1404 proteínas diferentes pertencentes à *Porphyromonas gingivalis*. No entanto o *OralInt v2.0* analisou 2301 proteínas pertencentes à espécie. Esta diferença mostra que nos resultados selecionados do *OralInt v2.0*, 897 proteínas pertencentes à espécie *Porphyromonas gingivalis* não apresentam nenhuma previsão de interação no conjunto de dados selecionados.

Do total de interações da rede do interactoma oral de *Porphyromonas gingivalis*, 862 interações são estabelecidas com proteínas sem atributo, que corresponde a 138 proteínas diferentes como explicado na secção 4.1.3.

Após análise topológica da rede do interactoma oral de *Porphyromonas gingivalis* recorrendo ao *Cytoscape* foi possível identificar quais as proteínas *hub* (Tabela 6).

Tabela 6- Proteínas *hub* da rede do interactoma oral de *Porphyromonas gingivalis*.

Código Uniprot	Nome da proteína	Espécie	Degree
Q9S3R9	<i>Membrane protein 41</i>	<i>Porphyromonas gingivalis</i>	3182
Q60FT2	<i>Putative uncharacterized protein pg0695</i>	<i>Porphyromonas gingivalis</i>	3181
B2RIQ3	<i>Membrane protein 41</i>	<i>Porphyromonas gingivalis</i>	3181
Q7MTP8	<i>Transporter Protein</i>	<i>Porphyromonas gingivalis</i>	1207
P60709	<i>Actin, cytoplasmic 1</i>	<i>Homo sapiens</i>	1128
Q536K6	<i>Beta actin</i>	<i>Homo sapiens</i>	1120

Analisando a tabela 6 podemos ver que as proteínas *hub* do interactoma oral de *Porphyromonas gingivalis* são comuns às proteínas *hub* da rede total do interactoma oral analisadas na seção 4.1.4. Este resultado reflete o elevado número de interações previstas para as actinas e assim é natural que estas estabeleçam muitas interações com as proteínas pertencentes também à *Porphyromonas gingivalis*.

A única proteína que não é comum nas duas redes é a *Putative uncharacterized protein pg0695* (Q60FT2), que apresenta 3181 interações na rede do interactoma oral de *Porphyromonas gingivalis*.

Dado ser uma proteína não caracterizada não se consegue perceber quais as funções biológicas que apresenta e como tal recorreu-se ao BLAST (*Basic Local Alignment Search Tool Blast*) para encontrar semelhança com outras proteínas, o que permite inferir relações funcionais bem como identificar a família de proteínas a que pertence.

Com o recurso ao BLAST foi possível averiguar que a proteína apresenta uma sequência de aminoácidos bastante semelhante a outras proteínas membranares da espécie *Porphyromonas gingivalis*. Uma das proteínas que apresenta identidade muito próximo dos 100% com esta, é a proteína *Membrane protein 41* (B2RIQ3).

Olhando para a tabela 6 podemos ver que na rede do interactoma oral de *Porphyromonas gingivalis* a *Membrane protein 41* (B2RIQ3) está presente e com o mesmo número de previsão de interações (3181).

Perante o resultado e após verificar as interações previstas para as duas proteínas, observou-se que o *OrallInt v2.0* previu as mesmas interações para ambas as proteínas e a realidade as 3 primeiras proteínas da tabela podem ser consideradas a mesma.

Deste resultado pode ver-se que o *OralInt v2.0* não fez distinção das proteínas, mas sim apesar de serem duas entradas diferentes assumiu como sendo a mesma proteína, o que levou a que calculasse as mesmas previsões de interação. Este resultado justifica-se pelo facto de as proteínas apresentarem sequências de aminoácidos iguais como constatado no BLAST e uma vez que o *OralInt v2.0* fez a previsão tendo por base as propriedades físico-químicas das proteínas, a elevada semelhança da sequência entre as proteínas fez com que os resultados obtidos fossem os mesmos.

Desta análise podemos ainda perceber que as três proteínas (Q9S3R9, Q60FT2 e B2RIQ3) com maior valor de *degree* na rede do interactoma oral de *Porphyromonas gingivalis*, que estabelecem 9544 interações nesta rede, apesar de três entradas diferentes apenas estamos perante uma proteína única. Deste resultado importa realçar duas situações diferentes: a primeira é que o facto de duas das proteínas não estarem revistas na base de dados Uniprot levou a uma duplicação da informação que certamente em fases posteriores de revisão levarão à eliminação das entradas exatamente com a mesma sequência. Por outro lado mesmo quando a revisão no Uniprot não foi feita seria útil o *OralInt v2.0* conseguir detetar que proteínas com a mesma sequência e pertencentes ao mesmo organismo deverão ser consideradas como repetição de informação, sendo apresentados nos resultados apenas uma delas.

Dado que as proteínas *hub* da rede do interactoma oral de *Porphyromonas gingivalis* não apresentam proteínas diferentes daquelas já discutidas anteriormente, decidiu-se olhar para o valor de *degree* das proteínas pertencentes à espécie *Porphyromonas gingivalis* para ver quais as proteínas que apresentam um valor de *degree* mais elevado (Tabela 7).

Tabela 7- Proteínas de *Porphyromonas gingivalis* com maior valor de *degree* na rede do interactoma oral de *Porphyromonas gingivalis*.

Código Uniprot	Nome da proteína	Valor de <i>degree</i>
<u>Q9S3R9</u>	<i>Membrane protein 41</i>	3182
<u>Q60FT2</u>	<i>Putative uncharacterized protein pg0695</i>	3181
<u>B2RIQ3</u>	<i>Membrane protein 41</i>	3181
<u>Q7MTP8</u>	<i>Transporter Protein</i>	1207
<u>Q7MWL3</u>	<i>30S ribosomal protein</i>	717
<u>Q7MVQ7</u>	<i>50S ribosomal protein</i>	663
<u>Q7MX16</u>	<i>Uncharacterized protein</i>	602
<u>Q7MXJ7</u>	<i>RagA protein</i>	552
<u>Q7MUM7</u>	<i>Conjugation transposon protein</i>	439
<u>Q9ZA60</u>	<i>Receptor antigen (RagA)</i>	431
<u>B2RLP6</u>	<i>1,4-alpha-glucan branching enzyme</i>	373
<u>B2RJU6</u>	<i>NADPH-NAD transhydrogenase beta subunit</i>	345
<u>Q7MU01</u>	<i>Transcription-repair-coupling factor</i>	326

Após análise da tabela 7 pode ver-se a presença de proteínas diferentes daquelas discutidas anteriormente (proteínas membranares). Podemos constatar no entanto, que as proteínas membranares associadas ao transporte apresentam um elevado número de previsão de interações. Este resultado é consistente com o facto das proteínas presentes nas membranas celulares, estarem expostas a uma grande diversidade de proteínas, quer de outros microrganismos, quer do hospedeiro. Como é através das proteínas de transporte que as moléculas essenciais ao funcionamento celular atravessam a membrana, é importante do ponto de vista do funcionamento celular que estabeleçam um elevado número de interações com outras proteínas. Assim os resultados obtidos do *OralInt v2.0* refletem estes dados, uma vez que prevê um elevado número de previsões para as proteínas membranares.

Apesar de analisarmos a rede de interações exclusivas com proteínas de *Porphyromonas gingivalis* ainda assim não figuram entre as proteínas com mais interações previstas pelo *OralInt v2.0* alguns fatores de virulência bem conhecidos e descritos para *Porphyromonas gingivalis* como as *cysteine proteinases (gingipains)* (45) e a *FimA* (46). Uma vez que é conhecido na comunidade científica que estas duas proteínas, pelas ações que têm durante a colonização e infeção do epitélio oral, terão de estabelecer interações com outras proteínas, decidimos procura-las no interactoma específico da *Porphyromonas gingivalis*.

As *gingipains* são proteínas proteolíticas produzidas pela *Porphyromonas gingivalis* que degradam o tecido periodontal e ativam a resposta imunológica do hospedeiro, são fatores de virulência fundamentais na periodontopatogênese (47). As *gingipains* têm ação de proteólise sob um vasto número de proteínas do hospedeiro, a bactéria aproveita os metabólitos resultantes dessa proteólise para o crescimento da própria célula e é também essa proteólise que conduz à desregulação das reações inflamatórias do hospedeiro (48).

Existem já descritas na literatura interações entre *gingipains* e outras proteínas como: *toll like receptors* (TLR) que estão presentes nas células CD14+, CD36+ e CD55+, recetores de complemento 3 (45) e também com o fibrinogénio, a fibronectina, a laminina e o colagénio tipo V (48). Algumas proteínas do hospedeiro são capazes de inibir a ação proteolítica das *gingipains* como a *lactoferrin* e a *histatin* (48).

Por outro lado, as proteínas *FimA* de *Porphyromonas gingivalis* contribuem para a colonização da espécie microbiana, a formação do biofilme oral, a invasão celular, a reabsorção óssea e a ativação de mecanismos do sistema imunológico do hospedeiro (49). Estas moléculas fimbriares são muitas vezes as que entram primeiro em contacto com as superfícies colonizadas pela bactéria desempenhando por isso um papel fundamental também no estabelecimento do biofilme.

Após análise da tabela 1 apenas os códigos B2RLK2, Q9ZNF9, P28784, P72197, P72194 e P95493 foram encontrados no *OralOme* como *gingipains*. Dentro da rede de interactoma de *Porphyromonas gingivalis* foi pesquisar-se as interações previstas para as duas proteínas.

A proteína *PorT* (Q9ZNF9), não apresenta nenhuma previsão de interação no conjunto de resultados selecionados do *OralInt v2.0*. Já a proteína *Lys-gingipain* (B2RLK2) apresenta 3 interações previstas. Interage com duas proteínas humanas que são: a *ADP-ribosylation factor 3* (P61204) e *ADP-ribosylation factor 1* (P84077) e uma interação com a *Cell wall surface anchor family protein* (Q82ZI0) pertencente à espécie *Enterococcus faecalis*. Nenhuma destas interações está descrita na literatura, portanto resta agora olhar para elas e perceber se poderão ser interações com potencial interesse. A anotação das interações pode ser observada na tabela 8.

Tabela 8- Relevância das interações estabelecidas pela proteína *Lys-gingipain* (B2RLK2) produzida pela *Porphyromonas gingivalis*.

Código Uniprot	Proteína com a qual interatua Espécie	Relevância da Interação
<u>P61204</u>	<i>ADP-ribosylation factor 3</i> (<i>Homo sapiens</i>)	As proteínas <i>ADP-ribosylation factor</i> têm função de transcrição, reparação do DNA, e serve de mediadora. São proteínas envolvidas em processos intermédios (50).
<u>P84077</u>	<i>ADP-ribosylation factor 1</i> (<i>Homo sapiens</i>)	
<u>Q82ZI0</u>	<i>Cell wall surface anchor family protein</i> (<i>Enterococcus faecalis</i>)	A interação era benéfica para o hospedeiro uma vez que a <i>Lys-gingipain</i> deixava de estar disponível para interagir com o hospedeiro e deixava a <i>Porphyromonas</i> com o seu mecanismo de virulência com menos agressividade para o hospedeiro. No entanto apesar do <i>Enterococcus</i> colonizar a cavidade oral ele coloniza mais os canais radiculares dos dentes (51) e a <i>Porphyromonas</i> no sulco gengival (52).

A proteína *Lys-gingipain* (P72197) apresenta 3 interações previstas com proteínas humanas, uma das interações é com a proteína *ADP-ribosylation 3* (P61204) que já foi abordada anteriormente. E apresenta duas interações com histonas que a anotação dessas interações pode ser consultada na tabela 9.

Tabela 9- Relevância das interações estabelecidas pela proteína *Lys-gingipain* (P72197) produzida pela *Porphyromonas gingivalis*.

Código Uniprot	Proteína com a qual interatua	Relevância da Interação
<u>Q71DI3</u>	<i>Histone H3.2</i>	As histonas são proteínas essenciais no nucleossoma. Desempenham um papel fundamental na regulação da transcrição, no reparo do DNA e na estabilidade cromossômica (53). Estudos recentes (54) referem propriedades antimicrobianas às histonas contra bactérias gram-positivas e gram-negativas. A previsão de interação com as histonas iriam fazer com que a <i>gingipains</i> perdesse a sua capacidade virulenta.
<u>Q16695</u>	<i>Histone H3.1t</i>	

A proteína *Lys-gingipain 381* (P72194) apresenta 7 interações previstas das quais 4 interações já foram discutidas anteriormente no presente trabalho (Q71DI3, Q16695, P61204 e P84077). A anotação das restantes 3 interações pode ser consultada na tabela 10.

Tabela 10- Relevância das interações estabelecidas pela proteína *Lys-gingipain 381* (P72194) produzida pela *Porphyromonas gingivalis*.

Código Uniprot	Proteína com a qual interatua Espécie	Relevância da Interação
<u>Q73QV2</u>	<i>Ribonucleoside-diphosphate reductase</i> (<i>Treponema denticola</i>)	A proteína têm função na replicação de DNA fornecendo percursos para a síntese do mesmo. E participa na síntese de desoxirribonucleicos (55).
<u>Q839G9</u>	<i>Elongation factor</i> (<i>Enterococcus faecalis</i>)	Proteína dependente de GTP, envolvida na formação da cadeia de peptídeos. É um elemento comum e altamente conservado no genoma (56).
<u>Q82ZJ0</u>	<i>Peptide deformylase</i> (<i>Enterococcus faecalis</i>)	Proteína envolvida na translocação, mas atua na remoção do grupo <i>formyl</i> nas proteínas recente sintetizadas (57).

De acordo com as interações apresentadas para a proteína *Lys-gingipain 381* (P72194) caso estas ocorram, podem comprometer o correto funcionamento celular das espécies (*Treponema denticola* e *Enterococcus faecalis*), uma vez que são estabelecidas interações com proteínas responsáveis pela regulação do processo de replicação bem como da formação de novas proteínas. Por outro lado, a interação da *Lys-gingipain* com proteínas microbianas faz com que esta não esteja disponível para interagir com o hospedeiro e fica assim a *Porphyromonas gingivalis* com menos um fator de virulência.

A proteína *gingipain R2* (P95493) apresenta 67 interações previstas pelo *OralInt v2.0* das quais 10 são estabelecidas com proteínas que foram eliminadas do Uniprot. A distribuição do número das proteínas (57) por espécie que interagem com a proteína *gingipain R2* pode ser visto na figura 12.



Figura 12- Número de proteínas distribuídas por espécie que interagem com a proteína *gingipain R2* (P95493) previstas pelo *OralInt v2.0*.

Dado ser uma proteína microbiana utilizada como fator de virulência por parte de *Porphyromonas gingivalis* é natural que estabeleça interações com proteínas do hospedeiro que representam 22/57 previstas pelo *OralInt v2.0*. Desta interações estabelecidas importa olhar para a *fibromodulin* (Q06828), é uma proteína que se encontra na matriz extracelular e que participa na formação das fibras de colagénio (58). Pensando no tecido gengival e na constituição do ligamento periodontal ele é rico em colagénio e como tal a proteína vai estar presente na matriz extracelular para formação das fibras, mas caso a interação com a *gingipain* ocorra esta proteína permanece na matriz extracelular mas não se liga para formar fibras e a presença de agentes patogénicos pode levar ao estabelecimento da doença periodontal. Um estudo (59) em tecido gengival humano, detetou uma maior quantidade da proteína *fibromodulin* no caso de tecido gengival inflamado associado (gingivite e periodontite) do que no caso de tecido gengival são. O que nos permite inferir que no caso do tecido são a proteína estava ligada ao colagénio a desempenhar a sua função. Uma interação com bastante interesse na medicina dentária.

Outra interação com potencial interesse é a estabelecida com a *Tax1-binding protein 1* (Q86VP1), esta proteína inibe a indução do fator TNF (*Tumor necrosis factor*). O TNF é composto por um grupo de citocinas capazes de provocar apoptose em células tumorais. Desempenha um papel importante na regulação do ciclo celular (60) e apresenta função pró-inflamatória. Se a proteína *Tax1-binding protein 1* (Q86VP1),

estiver a interagir com outras proteínas o fator TNF deixa de ter uma proteína reguladora negativa, e fica o fator TNF ativo para desempenhar as suas funções de controlo na deteção de células anormais, bem como ativar a resposta inflamatória do hospedeiro.

Da análise da tabela 2 apenas constam no *OralOme* os códigos B2RH54 e P59914 pertencentes a *FimA* da *Porphyromonas gingivalis*. Dentro da rede de interactoma da *Porphyromonas gingivalis* foi pesquisar-se as interações previstas para as duas proteínas.

A *Major fimbrial subunit protein type-1* (B2RH54) apresenta nos resultados quatro interações previstas, com proteínas de origem humanas e são elas: actin, cytoplasmic 1 (P60709), *beta actin* (Q53GK6), *actin* (Q9BYX7) e a *protein ribossomal 60S* (P62987). Olhando para estas proteínas elas foram já analisadas na secção 4.1.4 uma vez que se tratam das proteínas *hubs* da rede do interactoma oral. Este resultado justifica-se dado o número de interações previstas com elas de facto interagem com muitas das proteínas que constam na sub-rede criada.

A *Major fimbrial subunit protein* (P59914) na rede do interactoma da *Porphyromonas gingivalis* apresenta quatro interações previstas, duas delas pertencem à classe das actinas (P60709 e Q53GK6) que representam duas proteínas *hubs* da rede total do interactoma da cavidade oral em estudo e as outras *NADH dehydrogenase Ialpha subcomplex subunit 13* (Q9P0J0), *caulmodulin* (E7ETZ0). Nenhuma destas interações ainda foi descrita na literatura. A anotação referente a estas possíveis propostas de interações é discutida na tabela 11.

Tabela 11- Relevância das interações estabelecidas pela proteína *Major fimbrial subunit protein* (P59914) produzida pela *Porphyromonas gingivalis*.

Códigos <i>Uniprot</i>	Proteína com a qual interatua Espécie	Relevância da interação
<u>Q9P0J0</u>	<i>NADH dehydrogenase 1alpha subcomplex subunit 13</i> (<i>Homo sapiens</i>)	Proteína com função na cadeia respiratória mitocondrial (61). Pode desempenhar um na resposta das mucosas mediadas por CARD15 e regular a resposta das células epiteliais intestinais aos microrganismos (62).
<u>E7ETZ0</u>	<i>Caulmodulin</i> (<i>Homo sapiens</i>)	Atua como mediador de processos dependentes do cálcio e regula a concentração intra-celular do cálcio, funciona como uma enzima chave no sistema de transporte de cálcio. Apresenta funções celulares que incluem a sinalização, o metabolismo, a regulação do citoesqueleto e a regulação dos canais iônicos (63). Mas é uma proteína com atividade bastante importante no cérebro (64).

Após análise das interações previstas com proteínas pertencentes a dois fatores de virulência de *Porphyromonas gingivalis*, o *OralInt v2.0* no intervalo em estudo não detetou previsões de interações descritas na literatura o que dá um indicador negativo ao seu desempenho, isto porque, não está a reconhecer pares de interações positivos. Tendo em conta os resultados obtidos nas 200000 interações importa perceber se os pares de interações positivos (pares de interação com significado biológico), não serão apresentados pelo *OralInt v2.0* mas com um *score* menor. Isto porque as interações com *score* mais elevado parecem ser, pares de interação com maior compatibilidade nas sequências das proteínas e não pares de interação com significado biológico. Assim seria importante relacionar o *score* que o *OralInt v2.0* atribui às interações com o seu significado biológico e desta forma, não selecionar as interações com melhor *score* mas sim estabelecer-se o intervalo de valores de *score* onde fosse possível identificar os pares de interações positivos e novas propostas de interações com significado biológico.

4.2.2 Interactoma oral de *Streptococcus mutans*

Dado que na secção 4.2 foi verificado que existiam interações estabelecidas entre proteínas produzidas por espécies cariogénicas e no sentido de verificar se algumas das interações descritas na literatura eram reconhecidas pelo *OralInt v2.0*, seleccionou-se um cariogénico bem estudado, o *Streptococcus mutans*. Foi extraído da rede total do interactoma oral, apenas o grupo de interações em que estão envolvidas proteínas de *Streptococcus mutans*.

A rede do interactoma oral de *Streptococcus mutans* apresenta 28519 interações previstas. Deste total de interações, 450 interações são estabelecidas com proteínas sem atributo que corresponde a 142 proteínas diferentes como explicado na secção 4.1.3.

Da rede do interactoma oral de *Streptococcus mutans* estão representadas 964 proteínas diferentes pertencentes ao cariogénico. No entanto das proteínas que o *OralInt v2.0* analisou estavam presentes apenas 1393 proteínas pertencentes à espécie *Streptococcus mutans*. Esta diferença significa que dos resultados seleccionados do *OralInt v2.0*, 429 proteínas pertencentes à espécie *Streptococcus mutans* não apresentam nenhuma previsão de interação com melhor nível de confiança.

Após análise topológica da rede do interactoma oral de *Streptococcus mutans* conseguiu-se extrair as proteínas *hub* desta rede (Tabela 12).

Tabela 12- Proteína *hub* da rede do interactoma oral de *Streptococcus mutans*.

Código Uniprot	Nome da proteína	Valor de <i>degree</i>
<u>Q8DUC4</u>	<i>Uncharacterized protein</i>	1418
<u>I6L910</u>	<i>Putative PTS system, enzyme IIB component</i>	1183

Olhando para a tabela 12 podemos ver que, a proteína que apresenta um maior valor de *degree* é uma proteína que ainda não está caracterizada. Este resultado traduz-se sob o ponto de vista biológico insatisfatório, porque uma proteína não caracterizada não diz nada sobre as suas funções biológicas.

Recorrendo ao BLAST desta proteína percebe-se que ela apresenta semelhanças bastante elevadas com muitas outras proteínas pertencentes ao género *Streptococcus*, mas que também não estão caracterizadas.

O facto do *OralInt v2.0* prever um elevado número de interações (1418) com uma proteína não caracterizada pode dever-se à sequência da proteína, poderá a sequência de aminoácidos apresenta locais de ligação a várias outras proteínas e assim explicar o resultado obtido. Mais uma vez importa que o *OralInt v2.0* olhe e dê mais ênfase a parâmetros biológicos do que aos parâmetros de sequência.

Dado que a proteína *hub* da rede do interactoma oral de *Streptococcus mutans* não apresenta valor biológico decidiu-se olha para o valor de *degree* das proteínas pertencentes à espécie *Streptococcus mutans* para ver quais as proteínas que apresentam um valor de *degree* mais elevado (Tabela 13).

Tabela 13- Proteína de *Streptococcus mutans* com maior valor de *degree* na rede do interactoma oral de *Streptococcus mutans*.

Código Uniprot	Nome da Proteína	Valor de <i>degree</i>
<u>Q8DUC4</u>	<i>Uncharacterized protein</i>	1418
<u>I6L910</u>	<i>Putative PTS system enzyme IIB component</i>	1138
<u>Q8DVR2</u>	<i>Putative permease</i>	985
<u>Q8DV97</u>	<i>Putative sodium-dependent transporter</i>	896
<u>P12655</u>	<i>PPTS system sucrose-specific EIIBC component</i>	697
<u>Q8DU71</u>	<i>Putative ABC transporter, ATP-binding protein</i>	685
<u>Q8DVV4</u>	<i>Elongation factor G</i>	542
<u>Q8DS85</u>	<i>Leucine-tRNA ligase</i>	520
<u>Q8DVZ5</u>	<i>DNA polymerase</i>	510
<u>Q8DUK4</u>	<i>Putative ABC transporter, ATP-binding protein</i>	506
<u>Q8DSV6</u>	<i>Putative D-3-Phosphoglycerate dehydrogenase</i>	387
<u>Q8DU41</u>	<i>DNA gyrase subunit A</i>	372
<u>Q8DV21</u>	<i>50S ribosomal protein</i>	350

Olhando para as proteínas apresenta na tabela 13 pode ver-se que estamos perante proteínas envolvidas em funções da própria célula, isto é, são proteínas envolvidas em processos intra-celulares. Como tal, são proteínas que permitem a sobrevivência celular e a sua interação com outras pode comprometer o funcionamento celular normal. Mas também novamente as proteínas transportadoras estão aqui representadas na proporção de 6/13.

Desta lista de proteínas, existe uma que deve ser analisada com maior detalhe a proteína *PPTS system sucrose-specific EIIBC component* (P12655).

A proteína *PPTS system sucrose-specific EIIBC component* (P12655) pertencente ao *Streptococcus mutans* e como se sabe este é um dos principais agentes responsáveis pela cárie dentária (65). A dieta do hospedeiro é importante na atividade cariogénica da bactéria uma vez que esta, apresenta mecanismos capaz de metabolizar

uma grande quantidade de hidratos de carbono provenientes da dieta do hospedeiro. Uma dieta rica em açúcares, implica que o produto final do metabolismo irá ser o ácido láctico que leva a uma diminuição do pH e por sua vez à desmineralização do esmalte (66). Um dos sistemas mais comuns utilizados é o sistema *phosphotransferase system* (PTS) que permite a ligação de moléculas de açúcar e estas passam através da membrana celular sendo então metabolizado dentro da célula (67). Desta forma, esta proteína pode dizer-se que é um fator de virulência que o *Streptococcus mutans* utiliza para estabelecer a cárie dentária. Se a proteína estabelecer interação com outras proteínas, deixa de estar disponível para estabelecer ligação com o açúcar proveniente da dieta, conseqüentemente deixa de metabolizar o açúcar que levará depois à diminuição do pH do meio salivar. Como resultado da diminuição do pH existirá a dissolução do esmalte dentário e a partir daqui a formação da lesão de cárie. Olhando para as suas funções e percebendo a forma como ela pode contribuir para o estabelecimento da cárie dentária é uma proteína com interesse na medicina dentária e pode ver-se que foram previstas 697 interações para esta proteína pelo *OralInt v2.0*.

Apesar de estarmos perante uma rede com menor número de interações estabelecidas a forma de ver se o *OralInt v2.0* detetou interações já descritas foi procurar a proteína conhecida já como fator de virulência do *Streptococcus mutans* a *glucosyltransferase* (68).

As *glucosyltransferase* são enzimas extracelulares que utilizam a sacarose proveniente da dieta do hospedeiro para produzir exopolissacarídeos que ficam a fazer parte do biofilme oral. Os exopolissacarídeos criam locais de ligação para outras espécies microbianas e assim forma-se uma comunidade estruturada e envolvida por uma matriz ricas nos exopolissacarídeos. Desta forma o *Streptococcus mutans* e outros microrganismos acidogénicos, produzem ácidos a partir dos produtos do metabolismo da sacarose, em consequência existe uma diminuição do pH que leva a uma dissolução do esmalte dentário em consequência leva ao aparecimento clínico da cárie (69).

Após análise da tabela 3 a proteína *glucosyltransferase-SI* (P13470) não consta na rede do interactoma oral de *Streptococcus mutans*, o que significa que apesar de a proteína ter sido utilizada como dado de *input* do *OralInt v2.0*, este não previu nenhuma interação no conjunto das 200000 interações selecionadas.

A proteína *glucosyltransferase-I* (P08987) apresenta 19 interações previstas das quais 3 com proteínas humanas e 16 proteínas microbianas (Figura13)

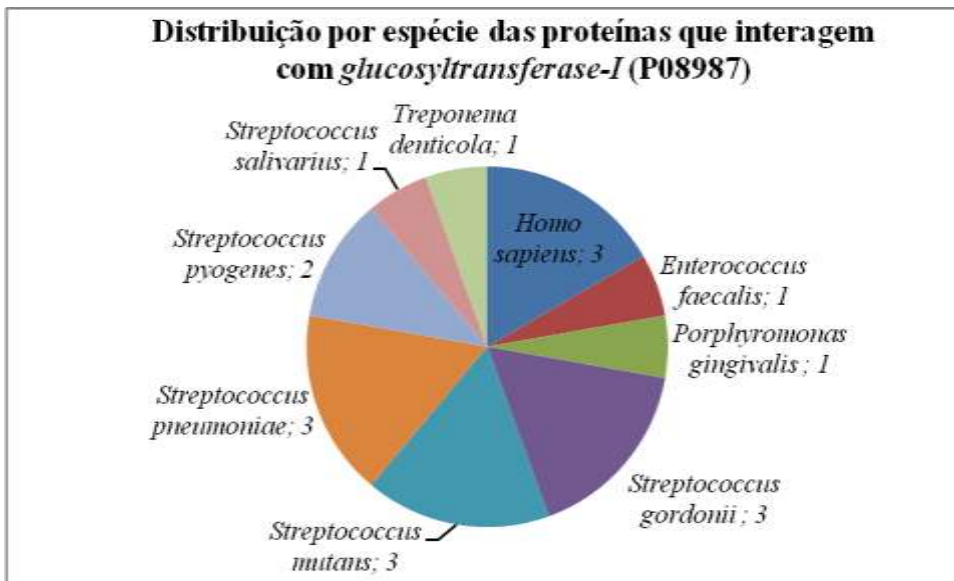


Figura 13- Representação das proteínas distribuídas por género que o interactoma oral do *Streptococcus mutans* apresenta como interações para a proteína *glucosyltransferase-I* (P08987)

Para a proteína *glucosyltransferase-I* (P08987) foi previsto um elevado número de interações com a proteína *Elongation factor* (8) pertencente a diferentes espécies (*Streptococcus mutans*, *Streptococcus salivarius*, *Streptococcus pyogenes*, *Streptococcus pneumoniae*, *Streptococcus gordonii* e *Enterococcus faecalis*). A proteína *Elongation factor* apresenta funções na síntese de novas proteínas controlando as mesmas para que estas sejam funcionais. A sua interação faz com que a sua função fique afetada e daí podem surgir novas proteínas com erros na sua sequência. A consequência pode ser grave para a célula se as proteínas com erros forem proteínas com funções celulares importantes.

A proteína *glucosyltransferase-S* (P49331) apresenta 32 interações previstas das quais 21 com proteínas humanas e 11 com proteínas microbianas (Figura 14)

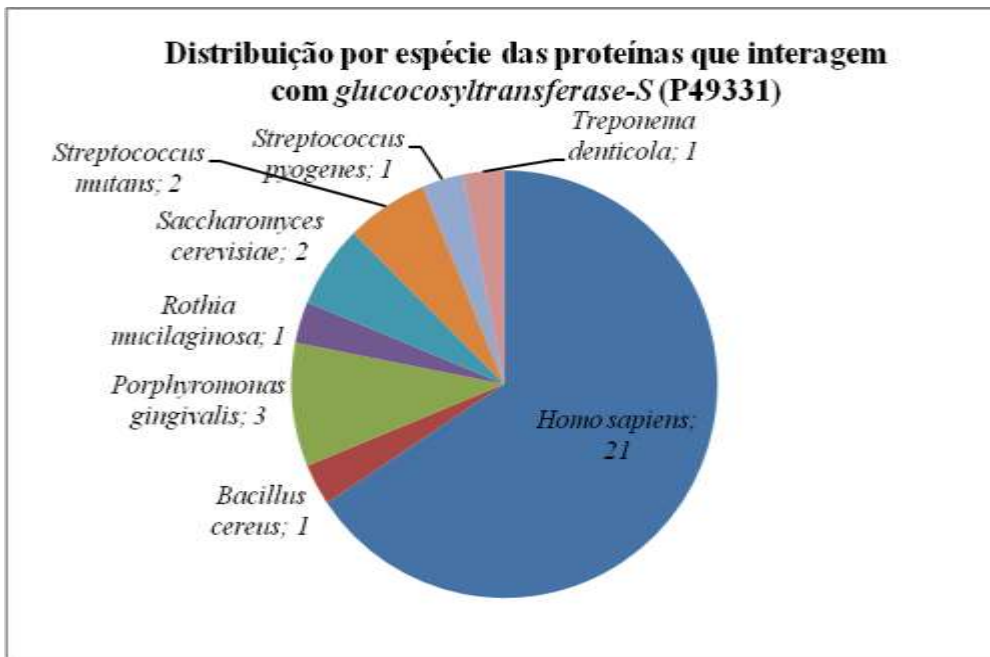


Figura 14- Representação das proteínas distribuídas por género que o interactoma oral do *Streptococcus mutans* apresenta como interações para a proteína *glucosyltransferase-S* (P49331).

Das 21 interações previstas com proteínas humanas 7 proteínas pertencem a proteínas *tubulin* são as proteínas major presentes nos microtúbulos, ligam-se ao GTP e permitem a transferência de energia ao longo dos microtúbulos.

A justificação para o elevado número de previsão de interações é a sequência de aminoácidos das proteínas apresentar uma semelhança elevada como constatado no BLAST. Mais uma vez, este resultado surge em consequência na forma como o *OralInt* v2.0 prevê as interações que é dando principal importância a parâmetros de sequência.

Após a análise dos resultados e apesar de um melhor desempenho do *OralInt* v2.0, pode ver-se que as proteínas com maior número de interações são proteínas transportadoras (proteínas microbianas) e proteínas com funções *housekeeping* (proteínas do hospedeiro). O que faz sentido dadas as suas funções e possuírem uma grande diversidade de domínios de adesão a outras moléculas. Apesar da rede, apresentar uma distribuição topológica idêntica a outras redes de PPIs a sua análise não se torna fácil, torna-se difícil extrair informação da rede dada a sua complexidade. A complexidade das redes PPIs é imensa e de facto os métodos computacionais vieram dar uma grande ajuda no que refere à previsão, mas os resultados positivos são escassos

(30, 70) o mesmo se verifica com o *OralInt v2.0* apesar de obter melhores resultados, os resultados obtidos não são ainda coesos para retirar pares de interações com significado biológico relevante quer para a explicação da formação do biofilme oral saudável, quer para as proteínas envolvidas nos mecanismos fisiopatológicos no estabelecimento de patologia oral pertencentes a microrganismos que habitam a cavidade oral.

5-CONCLUSÃO

A rede do interactoma oral obtida a partir dos resultados do *OralInt v2.0* veio a demonstrar um melhor desempenho do algoritmo no que se refere à previsão de interações entre proteínas do hospedeiro e proteínas microbianas face ao registado na versão anterior.

A análise topológica da rede revelou resultados idênticos aos que estão descritos para outras redes de PPI. O *OralInt v2.0* previu que a maioria das proteínas analisadas (74%) estabelecia um número reduzido de interações (1-25), e que um número muito restrito de proteínas apresentava uma previsão de interação muito superior (6%), sendo estas as proteínas *hub* da rede. No que respeita às proteínas *hub* da rede do interactoma oral, elas vieram a mostrar-se idênticas, aquelas que estão descritas em redes de PPIs humanas, uma vez que são proteína com funções estruturais de todas as células eucarióticas e não proteínas com funções específicas na cavidade oral.

A rede do interactoma oral revelou-se bastante grande, complexa e difícil de interpretar do ponto de vista biológico. A criação de sub-redes permitiu olhar para pares de interações específicas sob o ponto de vista biológico e permitiu ver que nas interações analisadas, não são encontradas interações já descritas na literatura, o que revelou, um aspeto negativo do desempenho do *OralInt v2.0*. Este é um dos aspetos a melhorar numa versão futura do *OralInt*.

Para além disto é ainda necessário afinar melhor o *OralInt v2.0*, no que respeita aos parâmetros de deteção das previsões. A utilização do código Uniprot como identificador no par de interações é muito útil mas faz com que por vezes as mesmas interações surjam muito repetidas. Ou seja, deveria também ser implementado numa nova versão do algoritmo uma forma de eliminar (ou descrever de uma forma mais concisa) interações estabelecidas entre proteínas com as mesmas sequências pertencentes aos mesmos organismos. Este aspeto é sobretudo importante nas proteínas bacterianas uma vez que estas estão menos revistas e constam nas bases de dados muitas sequências ou parciais ou totalmente repetidas.

Este tipo de estudo apesar de algumas limitações nos resultados e conclusões mais gerais é fundamental em áreas como a biologia oral em que é necessário olhar para a cavidade oral como um sistema multiorganismos. A maioria dos estudos de interatómica refere-se ao interactoma de um organismo e mesmo nos estudos em que há mais que um organismo são quase sempre estudados hospedeiros e microrganismos

patogénicos (*host-pathogen interactions*). Este é um dos únicos estudos em que um sistema é analisado na sua complexidade de organismos não necessariamente patogénicos e depois das afinações necessárias será fundamental no estudo de microbiomas complexos em que coexistem espécies microbianas, comensais, oportunistas e patogénicas a colonizarem um hospedeiro.

6-BIBLIOGRAFIA

1. Garnett JA MS. Interactions in Bacterial Biofilm Development: A Structural Perspective. *Current Protein & Peptide Science* 2012(13(8)):739-55.
2. Jorth P, Turner KH, Gumus P, Nizam N, Buduneli N, Whiteley M. Metatranscriptomics of the human oral microbiome during health and disease. *mBio*. 2014;5(2):e01012-14.
3. Heo SM, Ruhl S, Scannapieco FA. Implications of salivary protein binding to commensal and pathogenic bacteria. *Journal of oral biosciences / JAOB, Japanese Association for Oral Biology*. 2013;55(4):169-74.
4. Mei S, Zhu H. AdaBoost Based Multi-Instance Transfer Learning for Predicting Proteome-Wide Interactions between *Salmonella* and Human Proteins. *PloS one*. 2014;9(10):e110488.
5. Coelho ED AJ, Matos S, Pereira C, Rosa N, Correia MJ, et al. Computational prediction of the human-microbial oral interactome. *BMC systems biology*. 2014;8:24.
6. Morgan XC, Huttenhower C. Chapter 12: Human microbiome analysis. *PLoS computational biology*. 2012;8(12):e1002808.
7. Lamont RJ JH. *Oral microbiology at a glance*. John Wiley & Sons. 2010:p 29-54.
8. Pita PPC, Rodrigues JA, Ota-Tsuzuki C, Miato TF, Zenobio EG, Giro G, et al. Oral *Streptococci* Biofilm Formation on Different Implant Surface Topographies. *BioMed research international*. 2015;2015:159625.
9. Costalonga M, Herzberg MC. The oral microbiome and the immunobiology of periodontal disease and caries. *Immunology letters*. 2014;162(2 Pt A):22-38.
10. McLean JS. Advancements toward a systems level understanding of the human oral microbiome. *Frontiers in cellular and infection microbiology*. 2014;4:98.
11. Peterson SN, Meissner T, Su AI, Snesrud E, Ong AC, Schork NJ, et al. Functional expression of dental plaque microbiota. *Frontiers in cellular and infection microbiology*. 2014;4:108.
12. Zijng V, van Leeuwen MBM, Degener JE, Abbas F, Thurnheer T, Gmür R, et al. Oral Biofilm Architecture on Natural Teeth. *PloS one*. 2010;5(2):e9321.

13. Mahajan A, Singh B, Kashyap D, Kumar A, Mahajan P. Interspecies communication and periodontal disease. *TheScientificWorldJournal*. 2013;2013:765434.
14. Kiemer L, Cesareni G. Comparative interactomics: comparing apples and pears? *Trends in biotechnology*. 2007;25(10):448-54.
15. Droit A, Poirier GG, Hunter JM. Experimental and bioinformatic approaches for interrogating protein-protein interactions to determine protein function. *Journal of molecular endocrinology*. 2005;34(2):263-80.
16. Rao VS, Srinivas K, Sujini GN, Kumar GN. Protein-protein interaction detection: methods and analysis. *International journal of proteomics*. 2014;2014:147648.
17. Meyer K, Selbach M. Quantitative affinity purification mass spectrometry: a versatile technology to study protein-protein interactions. *Frontiers in Genetics*. 2015;6.
18. Nourani E, Khunjush F, Durmus S. Computational approaches for prediction of pathogen-host protein-protein interactions. *Frontiers in microbiology*. 2015;6:94.
19. Vengadesan K, Narayana SV. Structural biology of Gram-positive bacterial adhesins. *Protein science : a publication of the Protein Society*. 2011;20(5):759-72.
20. Rees MA, Kleifeld O, Crellin PK, Ho B, Stinear TP, Smith AI, et al. Proteomic characterization of a natural host-pathogen interaction: repertoire of in vivo expressed bacterial and host surface-associated proteins. *Journal of proteome research*. 2015;14(1):120-32.
21. Liu Y, Xie D, Han L, Bai H, Li F, Wang S, et al. EHFPI: a database and analysis resource of essential host factors for pathogenic infection. *Nucleic acids research*. 2015;43(Database issue):D946-55.
22. Qi Y, Noble W. Protein Interaction Networks: Protein Domain Interaction and Protein Function Prediction. In: Lu HH-S, Schölkopf B, Zhao H, editors. *Handbook of Statistical Bioinformatics*. Springer Handbooks of Computational Statistics: Springer Berlin Heidelberg; 2011. p. 427-59.
23. Smith DR. The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? *Briefings in functional genomics*. 2015.
24. Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, et al. The BioGRID interaction database: 2015 update. *Nucleic acids research*. 2015;43(Database issue):D470-8.

25. Zahiri J BJ, Masoudi-Nejad A. . Computational Prediction of Protein–Protein Interaction Networks: Algo-rithms and Resources. *Current Genomics* 2013;14(6):397-414.
26. Rosa N, Correia MJ, Arrais JP, Lopes P, Melo J, Oliveira JL, et al. From the salivary proteome to the OralOme: comprehensive molecular oral biology. *Archives of oral biology*. 2012;57(7):853-64.
27. Arrais JP, Rosa N, Melo J, Coelho ED, Amaral D, Correia MJ, et al. OralCard: A bioinformatic tool for the study of oral proteome. *Archives of oral biology*.58(7):762-72.
28. Doerks T, Copley RR, Schultz J, Ponting CP, Bork P. Systematic identification of novel protein domain families associated with nuclear functions. *Genome research*. 2002;12(1):47-56.
29. Petrey D, Honig B. Structural bioinformatics of the interactome. *Annual review of biophysics*. 2014;43:193-210.
30. Luo J, Qi Y. Identification of Essential Proteins Based on a New Combination of Local Interaction Density and Protein Complexes. *PloS one*. 2015;10(6):e0131418.
31. Luo J, Kuang L. A new method for predicting essential proteins based on dynamic network topology and complex information. *Computational biology and chemistry*. 2014;52:34-42.
32. Li M LY, Wang J, Wu FX, Pan Y. A topology potential-based method for identifying essencial proteins from PPI network *Computational Biology and Bioinformatics, IEEE/ACM Transactions*. 2015;12(2):372-83.
33. Zhang X, Xu J, Xiao WX. A new method for the discovery of essential proteins. *PloS one*. 2013;8(3):e58763.
34. Przulj N, Wigle DA, Jurisica I. Functional topology in a network of protein interactions. *Bioinformatics*. 2004;20(3):340-8.
35. ING C. PPPRINT Prediction of Protein-Protein Interaction [Mestrado]: Universidade de Coimbra; 2014.
36. UniProt C. UniProt: a hub for protein information. *Nucleic acids research*. 2015;43(Database issue):D204-12.
37. Leinonen R, Diez FG, Binns D, Fleischmann W, Lopez R, Apweiler R. UniProt archive. *Bioinformatics*. 2004;20(17):3236-7.
38. ANG S. Interatômica da cavidade oral [Mestrado]: Universidade Católica Portuguesa; 2014.

39. Murphy ACH, Young PW. The actinin family of actin cross-linking proteins – a genetic perspective. *Cell & Bioscience*. 2015;5:49.
40. Kowalewski JM, Shafqat-Abbasi H, Jafari-Mamaghani M, Endrias Ganebo B, Gong X, Stromblad S, et al. Disentangling Membrane Dynamics and Cell Migration; Differential Influences of F-actin and Cell-Matrix Adhesions. *PloS one*. 2015;10(8):e0135204.
41. Lima T, Auchincloss AH, Coudert E, Keller G, Michoud K, Rivoire C, et al. HAMAP: a database of completely sequenced microbial proteome sets and manually curated microbial protein families in UniProtKB/Swiss-Prot. *Nucleic acids research*. 2009;37(Database issue):D471-D8.
42. James MJ, Coulthurst SJ, Palmer T, Sargent F. Signal peptide etiquette during assembly of a complex respiratory enzyme. *Molecular microbiology*. 2013;90(2):400-14.
43. Silva N, Abusleme L, Bravo D, Dutzan N, Garcia-Sesnich J, Vernal R, et al. Host response mechanisms in periodontal diseases. *Journal of Applied Oral Science*. 2015;23(3):329-55.
44. Jordan F, Nguyen TP, Liu WC. Studying protein-protein interaction networks: a systems view on diseases. *Briefings in functional genomics*. 2012;11(6):497-504.
45. Jayaprakash K, Khalaf H, Bengtsson T. Gingipains from *Porphyromonas gingivalis* play a significant role in induction and regulation of CXCL8 in THP-1 cells. *BMC Microbiology*. 2014;14:193-.
46. Mantri CK, Chen C-H, Dong X, Goodwin JS, Pratap S, Paromov V, et al. Fimbriae-mediated outer membrane vesicle production and invasion of *Porphyromonas gingivalis*. *MicrobiologyOpen*. 2015;4(1):53-65.
47. Laheij AMGA, van Loveren C, Deng D, de Soet JJ. The impact of virulence factors of *Porphyromonas gingivalis* on wound healing in vitro. *Journal of Oral Microbiology*. 2015;7:10.3402/jom.v7.27543.
48. Olsen I, Potempa J. Strategies for the inhibition of gingipains for the potential treatment of periodontitis and associated systemic diseases. *Journal of Oral Microbiology*. 2014;6:10.3402/jom.v6.24800.
49. Khan SA, Kong EF, Meiller TF, Jabra-Rizk MA. Periodontal Diseases: Bug Induced, Host Promoted. *PLoS pathogens*. 2015;11(7):e1004952.
50. Morgan RK, Cohen MS. A Clickable Aminoxy Probe for Monitoring Cellular ADP-Ribosylation. *ACS chemical biology*. 2015;10(8):1778-84.

51. Fan W, Wu D, Ma T, Fan B. Ag-loaded mesoporous bioactive glasses against *Enterococcus faecalis* biofilm in root canal of human teeth. *Dental materials journal*. 2015;34(1):54-60.
52. Grenier D, Chen H, Ben Lagha A, Fournier-Larente J, Morin M-P. Dual Action of Myricetin on *Porphyromonas gingivalis* and the Inflammatory Response of Host Cells: A Promising Therapeutic Molecule for Periodontal Diseases. *PloS one*. 2015;10(6):e0131758.
53. Singh R, Harshman SW, Ruppert AS, Mortazavi A, Lucas DM, Thomas-Ahner JM, et al. Proteomic profiling identifies specific histone species associated with leukemic and cancer cells. *Clinical Proteomics*. 2015;12(1):22.
54. Hong SW, Seo DG, Baik JE, Cho K, Yun CH, Han SH. Differential profiles of salivary proteins with affinity to *Streptococcus mutans* lipoteichoic acid in caries-free and caries-positive human subjects. *Molecular oral microbiology*. 2014;29(5):208-18.
55. Truman AW, Kristjansdottir K, Wolfgeher D, Ricco N, Mayampurath A, Volchenbom SL, et al. Quantitative proteomics of the yeast Hsp70/Hsp90 interactomes during DNA damage reveals chaperone-dependent regulation of ribonucleotide reductase. *Journal of proteomics*. 2015;112:285-300.
56. Li X, Xing J, Li B, Wang P, Liu J. Use of tuf as a target for sequence-based identification of Gram-positive cocci of the genus *Enterococcus*, *Streptococcus*, coagulase-negative *Staphylococcus*, and *Lactococcus*. *Annals of Clinical Microbiology and Antimicrobials*. 2012;11:31-.
57. Cui P, Wang Y, Chu W, Guo X, Yang F, Yu M, et al. How water molecules affect the catalytic activity of hydrolases - A XANES study of the local structures of peptide deformylase. *Scientific Reports*. 2014;4:7453.
58. Zheng Z, Jian J, Velasco O, Hsu C-y, Zhang K, Levin A, et al. Fibromodulin Enhances Angiogenesis during Cutaneous Wound Healing. *Plastic and Reconstructive Surgery Global Open*. 2014;2(12):e275.
59. Qian H, Xiao Y, Bartold PM. Immunohistochemical localization and expression of fibromodulin in adult rat periodontium and inflamed human gingiva. *Oral Diseases*. 2004;10(4):233-9.
60. Journo C, Filipe J, About F, Chevalier SA, Afonso PV, Brady JN, et al. NRP/Optineurin Cooperates with TAX1BP1 to Potentiate the Activation of NF- κ B by Human T-Lymphotropic Virus Type 1 Tax Protein. *PLoS pathogens*. 2009;5(7):e1000521.

61. Cossard R, Esposito M, Sellem CH, Pitayu L, Vasnier C, Delahodde A, et al. *Caenorhabditis elegans* expressing the *Saccharomyces cerevisiae* NADH alternative dehydrogenase Ndi1p, as a tool to identify new genes involved in complex I related diseases. *Frontiers in Genetics*. 2015;6:206.
62. Lufei C, Ma J, Huang G, Zhang T, Novotny-Diermayr V, Ong CT, et al. GRIM-19, a death-regulatory gene product, suppresses Stat3 activity via functional interaction. *The EMBO Journal*. 2003;22(6):1325-35.
63. Haziza S, Magnani R, Lan D, Keinan O, Saada A, Hershkovitz E, et al. Calmodulin Methyltransferase Is Required for Growth, Muscle Strength, Somatosensory Development and Brain Function. *PLoS Genetics*. 2015;11(8):e1005388.
64. Timofeeva Y, Volynski KE. Calmodulin as a major calcium buffer shaping vesicular release and short-term synaptic plasticity: facilitation through buffer dislocation. *Frontiers in Cellular Neuroscience*. 2015;9:239.
65. Ajami B, Abolfathi G, Mahmoudi E, Mohammadzadeh Z. Evaluation of Salivary *Streptococcus mutans* and Dental Caries in Children with Heart Diseases. *Journal of Dental Research, Dental Clinics, Dental Prospects*. 2015;9(2):105-8.
66. Lynch DJ, Michalek SM, Zhu M, Drake D, Qian F, Banas JA. Cariogenicity of *Streptococcus mutans* Glucan-Binding Protein Deletion Mutants. *Oral health and dental management*. 2013;12(4):191-9.
67. Ajdic D, Chen Z. A novel phosphotransferase system of *Streptococcus mutans* is responsible for transport of carbohydrates with alpha-1,3 linkage. *Molecular oral microbiology*. 2013;28(2):114-28.
68. Ricker A, Vickerman M, Dongari-Bagtzoglou A. *Streptococcus gordonii* glucosyltransferase promotes biofilm interactions with *Candida albicans*. *Journal of Oral Microbiology*. 2014;6:10.3402/jom.v6.23419.
69. Falsetta ML, Klein MI, Colonne PM, Scott-Anne K, Gregoire S, Pai C-H, et al. Symbiotic Relationship between *Streptococcus mutans* and *Candida albicans* Synergizes Virulence of Plaque Biofilms In Vivo. *Infection and immunity*. 2014;82(5):1968-81.
70. Peng W WJ, Cheng Y, Lu Y, Wu F, Pan Y. UDoNC: An Algorithm for Identifying Essential Proteins Based on protein Domains and Protein-Protein Interaction Network. *Computational Biology and Bioinformatics, IEEE/ACM Transactions*. 2015;12(2):372-83.

