

Using Topic model and LSH for finding similar articles in PubMed

Peter Csiba, csiba4@uniba.sk, <http://www.csip.sk>

31-01-2014

1 Introduction

Task. Our aim was to find related articles in the PubMed article database. Pubmed is an open database of about 22 million science articles in medicine. We wanted to design something like a recommender system for science articles meaning that "if you found interesting this article, you should also look at". We want to apply the results for our project "sci-curve.com" to link similar articles and maybe visualizing this tree and clusters, i.e. topics.

Please note that almost all of the code was written by the author and is licenses under GPLv3.

Keywords. PubMed, Topic Model, LSH, LSI, gensim, python, c++, recommender system, kNN, cosine similarity.

Our approach. First of all, as the dataset is quite big, we selected 1,6M articles which have citations (that's about 5%). Then we parsed the articles from XML to following datasets:

- **corpus.csv** - tokenized abstracts
- **citations.csv** - one-to-many mapping from *one* article which cited *many*

Then we transformed the corpus to:

- **dictionary.pkl** - one-to-one mapping between token (from corpus) and it's unique integer identifier.
- **word_frequencies.csv** - for each word frequency of usage in all documents (from corpus).
- **docs_used.csv** - filtered and transformed tokens; for example number where omitted.

After the data was preprocessed we trained a topic model on *docs_used* with gensim using LSI:

- **lsi.pkl** - the trained model (it took about 16h).
- **topics.csv** - most significant tokens for each topic.
- **vectors.csv** - transformed *docs_used* to LSI space. These vectors should be compared with cosine similarity.

Finally, we developed two metrics based on LSI vectors and the citation graph to calculate the distance between two articles. To fasten the computation we used our LSH implementation in C++. Note that we first tried the bruteforce $O(N^2)$ kNN implementation to know if our metrics are working.

- **bf_dist.csv** - tuples (dist, id1, id2) of the closest articles. For each bucket in LSH we did $O(n^2)$ kNN, inserted all pairs to a global list and sorted. Note that only 2% of all were stored as there were over 1,000,000,000 pairs.
- **buckets.csv** - foreach bucket LSH the list of articles whic belong to them.

Note that we also had two smaller datasets of size 10,000 (small) and 100,000 (citation_small) which were used for testing purposes and preliminary results.

Also note that to download the PubMed dataset we used their open API. Pubmed allows to use its API to download the whole dataset.

Output. We compare the given cosine-similarity distance with exact citation walk distance on several hundred examples (note that it's computationally expensive to run BFS on such a graph). We also show example of few abstract (i.e. text) pairs with different distances.

To illustrate the dataset we include plots showing distributions of the data we used and transformed.

Please, find the attached source code in the zip file or check out <http://www.csip.sk/?p=1208>.

2 What we did

In this section we will specify our process. We will focus on the most significant machine learning parts and settings. We will skip data transformations.

2.1 Corpus generation

As mentioned before we transformed the corpus to:

- **dictionary.pkl** - one-to-one mapping between token (from corpus) and its unique integer identifier.
- **word_frequencies.csv** - for each word frequency of usage in all documents (from corpus).
- **docs_used.csv** - filtered and transformed tokens; for example number where omitted.

Most importantly:

```
tokens = [x.lower() for x in wordpunct_tokenize(abstract)]
doc_to_use = [w for w in tokens
if w not in skipwords and len(w) >= MIN_WORD_LENGTH and re.
    match(WORD_REGEX, w)]
#where
skipwords = set(list(stopwords) + list(sciencewords) + list(
    frequentwords))
#where
sciencewords = set(["abstract", "publisher", "welcome", "
    letters", "article", "editor", "articles", "published", "
    publish", "editing", "author", "authors", "original", "reply
    ", "previous", "publication", "publications", "appl", "opt",
    "appendix", "errors", "corrected"])
#where
frequentwords = [w for w,freq in frequencies.iteritems() if
    freq > 0.15]
#where
WORD_REGEX = r'^[a-zA-Z]+[a-zA-Z0-9]*$';
#skipword : see appendix A
```

Also, same abstracts, short abstracts and missing abstracts were filtered out. We also considered:

- adding POS-TAGS to words and taking only nouns so most of the comparison words will get out: "achieving better results" -> "results",
- or using only stem of words so "cells", "cellular", "cell" will get the same "cell" token.

We want to try this in future and compare the performance.

2.2 Topic model training

We just followed <http://radimrehurek.com/gensim/tutorial.html> to implement the following steps:

Note that it took 15h to train the final model. Before that we tried out smaller sample datasets and tweak the corpus (which words to include) and settings (e.g. how many topics).

```
tfidf = gensim.models.TfidfModel(corpus_memory_friendly)
lsi = gensim.models.lsimodel.LsiModel(corpus=tfidf[
    corpus_memory_friendly], id2word=dictionary, num_topics=
    NUM_TOPICS)
cPickle.dump(lsi, open("model/lsi_" + db_name + ".pkl", "w"))
```

The most challenging part here was to manage the memory as it was very inconvenient to keep all the corpus in memory. This version of LSI was iterative and for each iteration it took about 20,000 documents. About 10GB of RAM was used for the 1,6M articles with 150 topics.

2.3 k NN

This part was implemented in C++. We used hyperplane splitting as LSH functions, because they work well with cosine-similarity (see References: 2).

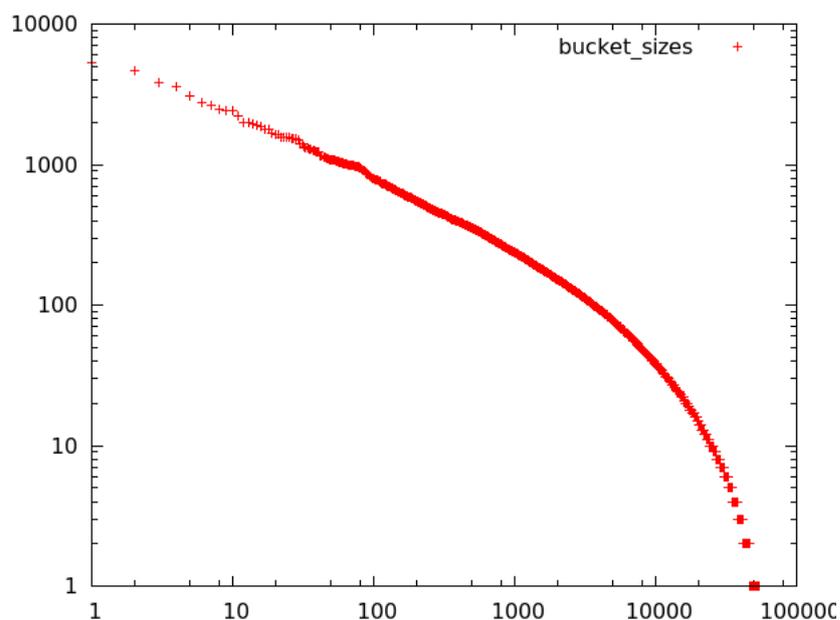
```
//Generating random hyperplanes
REP(i, LSH_BITS) REP(j, TOPIC_COUNT) lsh_fns[i][j] =
    rand_double(-1.0, 1.0);

//Hashing each point
REP(i, points.size()) {
    int hash = 0;
    REP(j, LSH_BITS) hash += (scalar(points[i], lsh_fns[j]) >
        0) * (1 << j);
```

LSH buckets size distribution. For each bucket we took each pair of points and calculated their cosine-similarity.

```
// foreach bucket
REP(i, buckets.size()){
    int n = buckets[i].size();

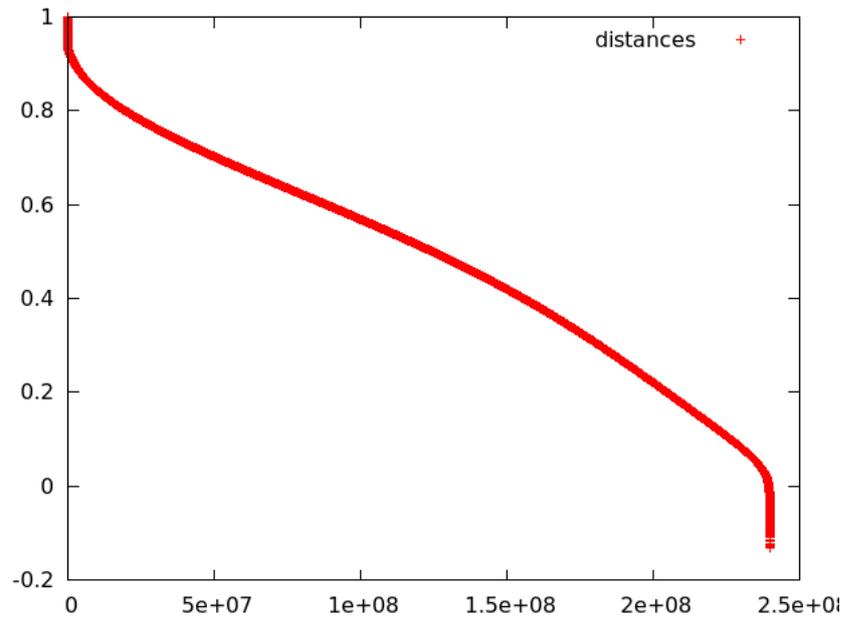
    // foreach unique bucket pair (distinct points)
    REP(a, n) REP(b, n) if(a < b){
        int p1 = buckets[i][a];
        int p2 = buckets[i][b];
        double cs = cos_sim(points[p1], points[p2]);
    }
}
```



We see that better selection of LSH functions could lead to more balanced bucket sizes. We guess that doing a "binary search like" or "selected random" search for each split could lead to more balanced bucket sizes. By "selected random" we mean generating N random hyperplanes and take the most balanced one.

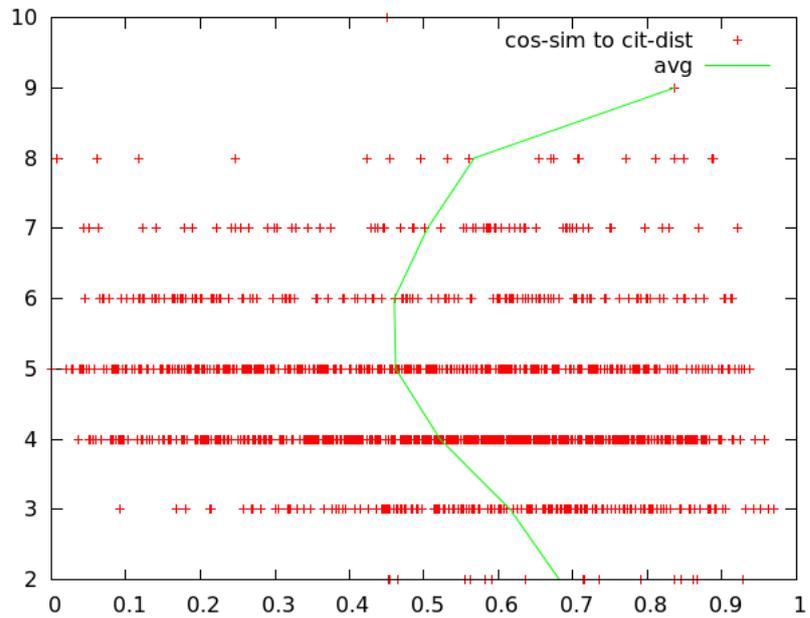
Getting closest buckets. This led to about 1,000,000,000 entries which we sorted and outputted 10,000,000 closest as well as each 100th to see the distribution of it.

Data is shown for $\cup_{b \in buckets} \{cosine_similarity(p_1, p_2) | p_1 < p_2 \in b\}$.



Comparing with citation walking distance. By *walking distance* we mean number of citation edges which need to be taken to reach from article A to article B .

We loaded the citation graph of 1.6M vertices and about 30M edges (as both ways). As the graph is quite big we implemented a version of BFS which tries only less then 2M edges (it could be configured). Then we selected 2408 articles uniformly from the cosine-distances computed from the last step. These distances were then compared with the citation distances as seen on the following plot:



3 References

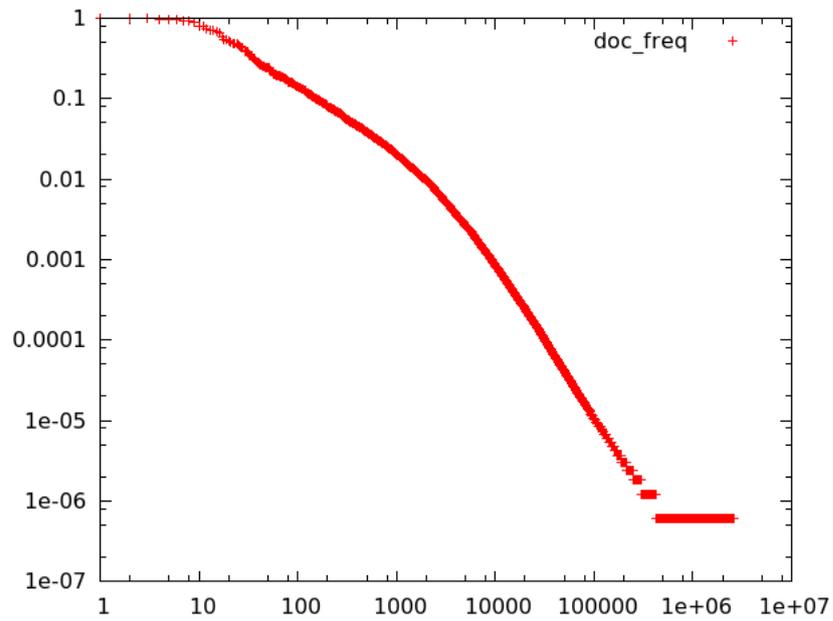
1. gensim: <http://radimrehurek.com/gensim/>
2. lsh: http://en.wikipedia.org/wiki/Locality-sensitive_hashing#Random_projection
3. pubmed: <http://www.ncbi.nlm.nih.gov/pubmed/>
4. scicurve: <http://scicurve.com/>
5. black swan (sponsor): <http://blackswanrational.com/>

For the code Python and C++ were used. For the plots GNUPlot and for typesetting LaTeX.

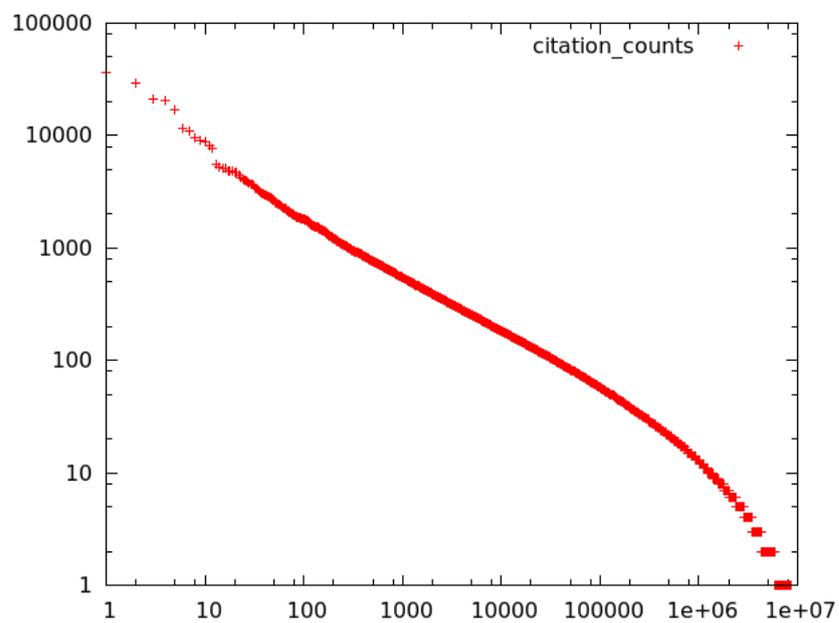
4 Appendix A - Dataset distributions

Stop words. "a", "about", "above", "after", "again", "against", "all", "am", "an", "and", "any", "are", "aren't", "as", "at", "be", "because", "been", "before", "being", "below", "between", "both", "but", "by", "can't", "cannot", "could", "couldn't", "did", "didn't", "do", "does", "doesn't", "doing", "don't", "down", "during", "each", "few", "for", "from", "further", "had", "hadn't", "has", "hasn't", "have", "haven't", "having", "he", "he'd", "he'll", "he's", "her", "here", "here's", "hers", "herself", "him", "himself", "his", "how", "how's", "i", "i'd", "i'll", "i'm", "i've", "if", "in", "into", "is", "isn't", "it", "it's", "its", "itself", "let's", "me", "more", "most", "mustn't", "my", "myself", "no", "nor", "not", "of", "off", "on", "once", "only", "or", "other", "ought", "our", "ours", "ourselves", "out", "over", "own", "same", "shan't", "she", "she'd", "she'll", "she's", "should", "shouldn't", "so", "some", "such", "than", "that", "that's", "the", "their", "theirs", "them", "themselves", "then", "there", "there's", "these", "they", "they'd", "they'll", "they're", "they've", "this", "those", "through", "to", "too", "under", "until", "up", "very", "was", "wasn't", "we", "we'd", "we'll", "we're", "we've", "were", "weren't", "what", "what's", "when", "when's", "where", "where's", "which", "while", "who", "who's", "whom", "why", "why's", "with", "won't", "would", "wouldn't", "you", "you'd", "you'll", "you're", "you've", "your", "yours", "yourself", "yourselves".

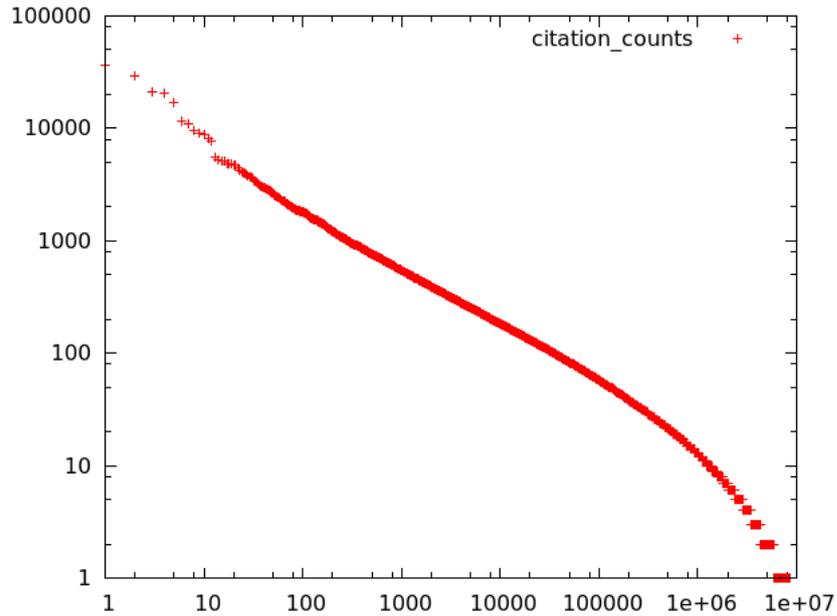
Document frequencies.



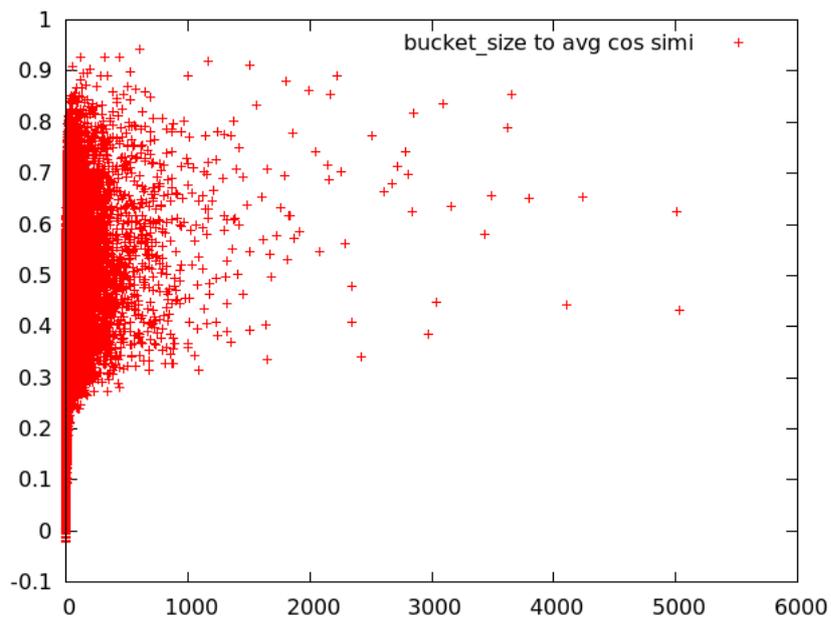
Citation counts.



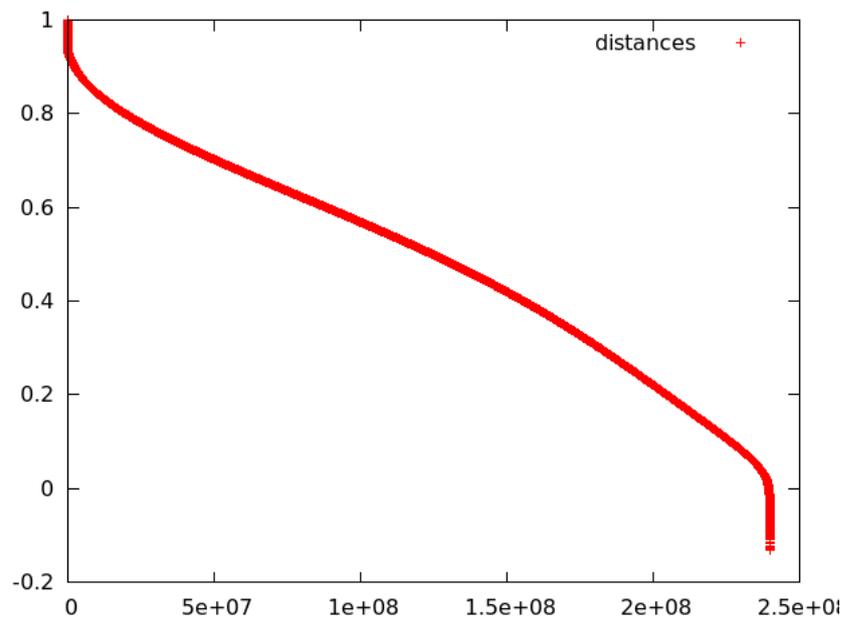
LSH buckets average cosine-similarity between pairs in bucket distribution. By "average cosine-similarity between pairs in bucket" we mean $avg(\{cosine_similarity(p_1, p_2) | p_1 < p_2 \in given_bucket\})$.



LSH buckets size (x) to average cosine-similarity between pairs (y).



Distance distribution of each pair in all buckets. Data is shown for $\cup_{b \in buckets} \{cosine_similarity(p_1, p_2 \in given_bucket)\}$.



Token frequencies in documents. (we skip stop words)

The following shows the measure $|\{word \in D\}|/|D|$

0.39340818557048773,results
0.2811760448033073,study
0.24174387666726485,cells
0.24098763179517527,may
0.23433867886421622,using
0.22367262519603848,cell
0.20359192306468835,protein
0.1979927100395108,can
0.1957101709533547,all
0.18752231972712766,data
0.18580936507241044,however
0.18539283019841823,associated
0.1853154051281805,one
0.18335637079286274,found
0.18192670786800766,patients
0.17958835070795923,methods
0.17268611576428447,analysis
0.1719682833301423,high
0.1609253076146056,increased
0.1577898923671485,specific
0.15654148813385774,activity
0.1552588728230201,studies
0.15149985565326052,compared
0.15009480069646552,expression
0.14841785770232402,gene
0.14615992658422797,human
0.1445111927242041,different
0.14210141243734722,showed
0.14185653314543248,role
0.1407005588409527,treatment
0.1383333923524437,significant
0.13761555991830154,three
0.1375273313498911,suggest
0.1350173186077652,significantly

5 Appendix B - Some of the topics

Here we analyse some of the topics with the most important tokens. We try to guess for what each topic stand.

Healthcare - healing people. 0.291 * patients, -0.164 * dna, 0.139 * health, 0.134 * risk, 0.126 * care, -0.118 * gene, -0.115 * binding, 0.113 * women, -0.106 * expression, -0.105 * rna, 0.104 * years, -0.100 * proteins, -0.098 * genes, 0.097 * age, 0.092 * children, 0.086 * cancer, 0.082 * patient, 0.082 * treatment, -0.080 * sequence, -0.078 * virus, -0.075 * alpha, -0.075 * beta, 0.074 * disease, 0.073 * clinical, 0.071 * group, 0.071 * mortality, -0.071 * transcription, 0.068 * year, 0.067 * months, -0.064 * membrane, -0.064 * mice, -0.062 * human, 0.062 * hospital, -0.062 * acid, -0.061 * activity, -0.061 * amino, -0.061 * receptor, -0.059 * sequences, 0.057 * diagnosis, 0.057 * medical, -0.057 * domain, 0.056 * pain, 0.056 * symptoms, 0.056 * men, -0.056 * activation, -0.056 * structure, 0.055 * cases, 0.054 * surgery, -0.053 * mrna, -0.053 * induced

Atoms and Molecules. 0.325 * mol, 0.275 * title, 0.266 * inter, 0.263 * hydrogen, 0.243 * ecules, 0.224 * ring, 0.217 * crystal, 0.214 * bonds, 0.203 * compound, 0.182 * dihedral, 0.171 * rings, 0.142 * angle, 0.140 * benzene, 0.133 * ecule, 0.129 * atoms, 0.117 * atom, 0.114 * actions, 0.111 * centroid, 0.108 * molecular, 0.102 * structure, 0.094 * bond, 0.077 * plane, 0.076 * chains, 0.073 * phenyl, 0.070 * weak, 0.070 * planar, 0.068 * axis, 0.066 * conformation, 0.066 * angles, 0.064 * linked, 0.064 * intra, -0.058 * mice, 0.057 * dimensional, 0.056 * unit, 0.056 * distorted, -0.056 * expression, 0.055 * along, 0.053 * group, 0.050 * inversion, 0.050 * asymmetric, 0.049 * stabilized, 0.047 * forming, 0.047 * water, 0.047 * adopts, 0.047 * dimers, 0.046 * link, -0.044 * virus, 0.044 * packing, 0.044 * pyridine, 0.043 * membered

Genetics. 0.219 * dna, -0.190 * mice, -0.171 * patients, 0.162 * health, 0.122 * genes, 0.116 * gene, 0.115 * sequence, 0.114 * rna, 0.106 * care, 0.102 * genome, 0.097 * sequences, -0.096 * induced, -0.088 * alpha, -0.086 * insulin, -0.083 * receptor, -0.079 * tumor, -0.077 * serum, -0.076 * rats, 0.076 * hiv, -0.075 * response, -0.074 * ifn, 0.073 * genetic, 0.071 * data, -0.071 * tnf, -0.071 * blood, -0.070 * activation, -0.070 * inflammatory, 0.070 * based, 0.070 * information, -0.069 * beta, -0.068 * treatment, -0.068 * levels, -0.067 * liver, -0.066 * anti, -0.066 * gamma, -0.066 * glucose, -0.065 * increased, -0.062 * immune, 0.062 * research, 0.061 * species, -0.061 * treated, 0.060 * analysis, -0.060 * plasma, -0.058 * dose, 0.057 * region, -0.057 * responses, 0.056 * use, 0.056 * method, -0.055 * muscle, -0.054 * lung

Viruses. -0.616 * hiv, -0.307 * virus, -0.238 * infection, -0.172 * infected, -0.170 * viral, -0.135 * cd4, -0.087 * hcv, -0.083 * replication, -0.080 * immune, -0.080 * immunodeficiency, -0.079 * rna, -0.078 * patients, -0.076 * viruses, -0.074 * antibody, -0.073 * cd8, -0.072 *

influenza, 0.070 * cancer, -0.070 * antigen, -0.070 * ifn, -0.068 * mice, -0.067 * antibodies, -0.065 * aids, -0.064 * vaccine, -0.060 * infections, -0.057 * dna, -0.056 * hepatitis, 0.053 * ca2, -0.050 * anti, -0.050 * strains, -0.050 * host, 0.050 * insulin, 0.050 * muscle, 0.049 * glucose, -0.047 * isolates, -0.045 * human, -0.044 * positive, -0.042 * hsv, -0.041 * antiretroviral, 0.041 * membrane, -0.040 * antigens, -0.040 * lymphocytes, 0.039 * channel, -0.039 * gamma, 0.039 * genes, 0.038 * neurons, -0.037 * hbv, 0.037 * growth, -0.037 * mol, 0.037 * brain, 0.036 * signaling

Cancer and hard to cure diseases. -0.412 * cancer, -0.267 * patients, -0.216 * breast, -0.169 * dna, -0.164 * tumor, 0.163 * hiv, 0.142 * health, -0.140 * gene, -0.132 * genes, -0.121 * expression, -0.098 * survival, -0.092 * tumors, 0.089 * children, 0.085 * ca2, -0.082 * prostate, 0.079 * membrane, 0.078 * glucose, 0.069 * insulin, -0.068 * carcinoma, -0.068 * p53, 0.068 * care, -0.067 * disease, -0.066 * cases, -0.064 * chemotherapy, -0.062 * cancers, -0.062 * lung, -0.058 * mutations, -0.057 * diagnosis, 0.057 * rats, -0.057 * rna, -0.056 * tumour, 0.055 * activity, -0.054 * metastasis, -0.054 * therapy, -0.054 * transcription, 0.054 * concentration, 0.052 * channel, -0.052 * genome, -0.051 * metastatic, -0.050 * surgery, -0.050 * colorectal, -0.050 * patient, 0.049 * muscle, -0.048 * promoter, 0.048 * channels, 0.048 * effects, 0.047 * social, 0.047 * release, 0.047 * concentrations, -0.046 * mir

6 Appendix C - Example of abstracts

In this section we show abstract of different cosine-similarity. The bold number shows the cosine-similarity of the following two abstracts. The two numbers in italic show the article id-s. We selected randomly one pair for cosine-similarities 0.0, 0.1, ..., 1.0.

Note that distance "1" means that the two abstracts are equivalent and distance "0" that are completely different. In both cases there could be precision loss due to doing almost 1,000 multiplications with summation and square root.

1 *21578896* in the title compound " " c (14) h (15) no (2) s " " the dihedral angle between the aromatic rings is 49 . 7 (1). in the crystal " " inversion dimers linked by pairs of n - h o hydrogen bonds occur .

21578810 in the title compound " " c (15) h (13) n (3) o (2) s (2 ") " the dihedral angle between the aromatic rings is 87 . 8 (2). in the crystal " " inversion dimers occur linked by pairs of n - h s hydrogen bonds .

0.900696 *23041316* hiv - 1 depends on many host factors for propagation . other host factors " " however " " antagonize hiv - 1 and may have profound effects on viral activation . curing hiv - 1 requires the reduction of latent viral reservoirs that remain in the face of antiretroviral therapy . using orthologous genetic screens " " we identified bromodomain containing 4 (brd4) as a negative regulator of hiv - 1 replication . antagonism of brd4 " " via rna interference or with a small molecule inhibitor " " jq1 " " both increased proviral transcriptional elongation and alleviated hiv - 1 latency in cell - line models . in multiple instances " " jq1 " " when used in combination with the nf - b activators prostratin or pha " " enhanced the in vitro reactivation of latent hiv - 1 in primary t cells . these data are consistent with a model wherein brd4 competes with the virus for hiv - 1 dependency factors (hdfs) and suggests that combinatorial therapies that activate hdfs and antagonize hiv - 1 competitive factors may be useful for curing hiv - 1 infection .

1401081 individuals infected with hiv may be asymptomatic for years before progressing to overt aids . since hiv can latently infect monocytoid cell lines " " we examined whether hiv latency occurs in monocytes in vivo . freshly isolated monocytes from asymptomatic seropositive individuals examined before and after culture were positive for hiv dna " " but not rna " " as measured by polymerase chain reaction " " showing that hiv latency occurs in monocytes in vivo . coculture of these latently infected monocytes with con a - activated t cells from hiv - negative normal donors stimulated 90 % of the patients ' samples and latently infected thp - 1 to produce infectious virus . neither con a " " resting t cells " " nor

t cell supernatants induced virus . plasma membranes from activated t cells stimulated hiv production ” ” suggesting cell contact induces factor (s) in monocytes to overcome latency . thus ” ” monocytes in aids patients harbor latent hiv inducible during an immune response ” ” leading to t cell infection and viral - induced pathology .

0.80003 20589091 recognizing syndromes which mimic als is crucial both to avoid giving this diagnosis erroneously and since there may be appropriate treatments . we report a 63 - year - old woman diagnosed with possible als five years ago based on upper and lower motor neuron signs with typical electrophysiology and normal cranial mri . at reassessment ” ” spinal mri revealed a cervicothoracic cyst with cord compression that was successfully treated neurosurgically . histopathology confirmed an arachnoid origin as suspected from mri . spinal cysts may mimic als and need to be thoroughly excluded by appropriate imaging .

22605960 we report a case of a 12 year old girl who presented with cord compression . imaging studies demonstrated an extradural spinal tumour in the lower thoracic and upper lumbar levels . histology confirmed the diagnosis of schwannoma while associated findings suggested the possibility of neurofibromatosis type i .

0.700191 7395865 four patients with hyperornithinemia and gyrate atrophy of the choroid and retina age described . in vivo response to vitamin b6 is documented in three of the four patients by significant reduction of fasting serum ornithine and increase of lysine after oral b6 supplementation . oral glucose tolerance testing in one patient resulted in marked changes in serum ornithine and lysine concentrations ” ” in addition to mild glucose intolerance . histochemical staining of punch muscle biopsies showed intracellular inclusions in type 2 muscle fibers . tubular aggregates ” ” approximately 60 nm in diameter and adjacent to the sarcoplasmic membrane ” ” were seen on electron microscopy . obligate heterozygotes had a mean serum ornithine slightly higher than normal ” ” but there was considerable overlap with the normal range . oral ornithine tolerance tests distinguished carriers from controls in only one of five cases . deficient activity of ornithine ketoacid aminotransferase (okt) in cultured skin fibroblasts was documented in all four patients . approximately half - normal levels were found in obligate heterozygotes

19244070 objective : this study examined quantitative heel ultrasonography (qus ”) ” 25 - hydroxyvitamin d (25 - ohd) levels ” ” and urine amino - terminal cross - linking telopeptide of type i collagen (ntx - i) levels in patients with a recent osteoporotic hip fracture to see whether they were clinically useful . methods : stiffness index (si) t scores from qus ” ” 25 - ohd levels ” ” and urine ntx - i levels were obtained in 53 female and 32 male patients with hip fractures . sixty - five female patients and 5 male patients attending our geriatric clinic were used for comparison . results : the si t scores of the hip fracture patients were less than those of the geriatric clinic patients . the difference was significant in female patients (p = . 0001) but not in male patients (p = . 1)

0.600245 23396624 purpose : to identify the vascular patterns found in superficial lymph nodes with histologically confirmed lymphomatous involvement and to determine their value in the sonographic diagnosis of lymphadenopathy . methods and materials : the study involved the prospective classification of vascular patterns observed during power doppler and / or color doppler studies of superficial lymph nodes scheduled for resection . forty patients (27 men and 13 women " " aged 22 - 84 years ; mean age : 58 years) with pathologically proven lymphoma were selected for this study (26 cervical " " 13 axillary and 1 inguinal). results : a longitudinal vessel with or without branches (pattern i) was found in 14 lymphomatous nodes . six contained short vessel segments distributed in the hilum area or centrally (pattern ii ") " five had multiple vessels " " partially branching " " entering the node in a few rows from its longitudinal side (pattern iii ") " seven presented multiple vessels that branched irregularly or chaotically with avascular areas (pattern iv ") " and eight had a peripheral vessel distribution (pattern v). therefore " " 50 % of the lymphomatous nodes had vascular patterns regarded as characteristic of reactive lymph nodes (patterns i and ii ") " and 37 . 5 % had patterns normally described in lymph nodes with metastatic involvement (patterns iv and v); other lymphomatous lymph nodes had ambiguous vascular patterns that have not been previously classified (pattern iii). conclusion : the angioarchitecture of superficial lymphomatous lymph nodes varies widely and is difficult to classify

1716455 the prognostic significance of argyrophilic nucleolar organiser regions (agnors) has been evaluated in biopsy specimens from 61 primary squamous and undifferentiated carcinomas of the pharynx prior to therapy . the univariate kaplan - meyer survival analysis showed a significant correlation between 3 - and 5 - year survival rates and the mean agnor number per tumour cell (p less than 0 . 001). no significant correlation was found between prognosis and patients age and sex " " tumour location " " clinical stage " " histologic grade " " extent of lymphocytic infiltration " " hmfg - 2 positivity of tumour cells and uchl1 " " ln2 " " mb2 positivity of infiltrating lymphocytes . there was no significant association between agnor counts and tumour histologic grade or clinical stage . multivariate survival analysis showed that only two variables were significantly correlated with prognosis : agnor counts (p less than 0 . 001) and the extent of lymphocytic infiltration (p less than 0 . 027). our results indicate the prognostic value of agnor counts and suggest the use of this method as a significant parameter in the pretherapeutic assessment of the aggressiveness of pharyngeal carcinomas .

0.500036 20498241 background and objectives : concentrated heparin solutions are instilled into the catheter lumens after each hemodialysis session to prevent catheter thrombosis . the heparin lock concentration at many centers has been decreased recently to reduce the risk of systemic bleeding and contain costs . however ” ” the effect of this change on catheter patency is unknown . we compared catheter patency between two heparin lock solutions : 1000 versus 5000 units / ml . design ” ” setting ” ” participants ” ” & measurements : with use of a prospective ” ” computerized ” ” vascular access database ” ” 105 patients with newly placed tunneled hemodialysis catheters ” ” including 58 patients receiving a 5000 units / ml heparin lock and 47 patients receiving a 1000 units / ml heparin lock ” ” were retrospectively identified . the primary endpoint was cumulative catheter patency and the secondary endpoint was frequency of thrombolytic instillation . results : cumulative catheter survival was similar in the two groups ” ” being 71 % versus 73 % at 120 days in the low - and high - concentration heparin lock groups (hazard ratio of catheter failure ” ” 0 . 97 ; 95 % confidence interval ” ” 0 . 45 to 2 . 09 ; p = 0 . 95) . the frequency of tissue plasminogen activator instillation was significantly greater in the low - concentration heparin group (hazard ratio ” ” 2 . 18 ; 95 % ci ” ” 1 . 26 to 3 . 86 ; p = 0 . 005) . no major bleeding complications were observed in either treatment group . the overall drug cost for maintaining catheter patency was 23 % lower with the low - concentration heparin lock (1418 *versus* 1917) to maintain catheter patency for 1000 days . conclusions : low - concentration heparin lock solutions do not decrease cumulative dialysis catheter patency ” ” but require a twofold increase in thrombolytic instillation to maintain long - term patency .

21555056 approximately 15 % of patients with sturge - weber syndrome demonstrate bilateral intracranial involvement ” ” and the prognosis of these patients is considered particularly unfavorable . we reviewed the clinical and neuroimaging features of patients with sturge - weber syndrome and bilateral intracranial involvement . seizure variables ” ” the presence of hemiparesis ” ” and the degree of developmental impairment at most recent follow - up were compared with imaging abnormalities . of 110 sturge - weber syndrome patients ” ” 14 demonstrated bilateral brain involvement ” ” with an asymmetric pattern on glucose metabolism positron emission tomography . although most patients manifested frequent seizures initially ” ” associated with frontal hypometabolism on positron emission tomography ” ” six (43 %) had achieved good seizure control during follow - up . bilateral frontal hypometabolism was associated with severe developmental impairment . two children with bitemporal hypometabolism exhibited autistic features . hemiparesis was associated with superior frontal (motor cortex) hypometabolism . three patients underwent resective surgery ” ” resulting in improved seizure control and developmental outcomes . the severity of neurologic complications and clinical course depend on the extent of cortical dysfunction in bilateral sturge - weber syndrome

0.400013 *21341961* health care providers can play a key role in influencing clients to initiate and maintain use of the female condom " " an underused method for hiv / sti and pregnancy prevention . in 2001 - 2002 " " based on semistructured interviews with 78 health care providers from four types of settings in new york city " " we found that most providers had seen the female condom " " but they had not used it and did not propose the method to clients . they lacked details about the method - when to insert it " " where it can be obtained " " and its cost . gender of provider " " provider level of training " " and setting appeared to influence their attitudes . unless and until provider training on the female condom is greatly improved " " broader acceptance of this significant public health contribution to preventing hiv / aids and unwanted pregnancy will not be achieved .

23008650 background : community pharmacists in saudi arabia very often make decisions that affect patient outcome . previous studies have indicated that they have access to limited sources of information . therefore " " structured continuing pharmaceutical education (cpe) is necessary to improve their standards and attitudes . aims : identify the most important topics for cpe as well as the most significant barriers to conducting cpe successfully . methods : a questionnaire was distributed to 120 pharmacists working in 88 community pharmacies in the eastern province of saudi arabia . the survey contained five sections : general background " " topics for cpe that could be of great interest to community pharmacists " " possible obstacles to attending cpe " " method of instruction " " and the most suitable time and day of the week for conducting cpe . results : one hundred and five (87 . 5 %) pharmacists answered the survey questionnaire . the rank order of the five most selected topics for cpe were : drug interaction (81 . 9 %) " drug use during pregnancy (77 . 1 %) " use of anti - microbial agents (62 . 5 %) " pharmaceutical ethics (53 . 3 %) " geriatric and pediatric pharmacology (45 . 7 %) . for pharmacists " " the most important obstacles to attending cpe were lack of time (96 . 2 %) " distance from practice (74 . 2 %) " and lack of programs or information about these programs (54 . 3 %) . interestingly " " 47 . 6 % of the pharmacists recommended credentialing cpe and stated that knowing the lecturer was not considered an important factor . conclusion and recommendations : the findings of this study demonstrated that pharmacists are willing to participate in cpe programs . however " " the working conditions of pharmacists would be a major barrier to their attendance

0.300394 7596814 in the macronucleus of the ciliate oxytricha nova " " telomeres end with single - stranded (t4g4) 2 dna bound to a heterodimeric telomere protein (alpha beta). both the alpha and beta subunits (alpha - tp and beta - tp) were phosphorylated in asynchronously growing oxytricha ; beta - tp was phosphorylated to a much higher degree . in vitro " " mouse cyclin - dependent kinases (cdks) phosphorylated beta - tp in a lysine - rich domain that is not required for specific dna binding but is implicated in higher order structure formation of telomeres . therefore " " phosphorylation of beta - tp could modulate a function of the telomere protein that is separate from specific dna binding . phosphoamino acid analysis revealed that the mouse cdks modify predominantly threonine residues in beta - tp " " consistent with the observation that beta - tp contains two consensus cdk recognition sequences containing threonine residues . in xenopus egg extracts that undergo cell cycling " " beta - tp was phosphorylated in m phase and dephosphorylated in interphase . this work provides the first direct evidence of phosphorylation at telomeres in any organism " " as well as indirect evidence for cell cycle regulation of telomere phosphorylation

6304689 a library of saccharomyces cerevisiae dna in the vector lambda charon 28 was probed for sequences complementary to cdna made from poly (a)+ rna isolated from the well - sporulating yeast strain ap1 a / alpha . the rna was isolated from cells that had been incubated 7 " " 9 " " 11 " " and 13 hr in sporulation medium . dna complementary to poly (a)+ rna from alpha / alpha (nonsporulating) ap1 was used as a control " " and 46 bacteriophage that gave a stronger response with a / alpha cdna than with alpha / alpha cdna were obtained in a screening of three yeast genomes worth of dna . two of the bacteriophage appeared to contain a / alpha - specific genes " " in that they hybridized to cdna from vegetative a / alpha rna . the rest appeared to correspond to a / alpha genes expressed preferentially during sporulation . restriction endonuclease analysis of four of the cloned sequences revealed a single major region of transcription in each ; these regions ranged in size from 2 . 5 to 4 . 0 kilobases . rna blot analysis showed that " " in three of the four cases " " transcripts of two different sizes were homologous to the cloned sequence . in all four cases " " the homologous transcripts appeared at about 7 hr and were decreasing in amount by 13 hr . these results provide evidence for transcriptional control of genes expressed during sporulation and for at least one group of genes that is turned on at about the time of meiosis i in sporulation .

0.200064 18489745 a 45 - year - old white male presented to our department with postoperative recurrence of gastrointestinal poorly differentiated neuroendocrine carcinoma manifesting as lymph node dissemination and a solitary implantation metastasis in the rectovesical pouch . following disease progression on chemotherapy ” ” the patient was treated with radiotherapy using either a conventional daily treatment or an accelerated hyperfractionated protocol to separate sites of disease progression . using serial ct scan assessment ” ” changes in cross - sectional area of the separately treated metastatic lesions were evaluated for determination of therapy response . the accelerated hyperfractionated radiotherapy appeared to limit the rate of tumor growth to a greater degree than the conventional fractionation schedule . of uttermost importance ” ” in this palliative setting ” ” the patient completed the intensified radiotherapy regimens with acceptable acute toxicity . given the proliferative capacity of poorly differentiated neuroendocrine carcinomas of the gastrointestinal tract ” ” radiotherapy may be a therapeutic supplement to chemotherapy ” ” which represents the main treatment option in this tumor entity . importantly ” ” tumors with a capacity for rapid proliferation and regeneration may be particularly sensitive to the use of intensified fractionation protocols in clinical radiotherapy .

18751768 background : the risk of radiographic knee degeneration after partial or total meniscectomy is well documented ” ” but no prior study has employed cartilage - sensitive mri technology to assess degenerative changes after meniscectomy . hypothesis : arthroscopic partial meniscectomy results in early articular cartilage wear and subchondral bony degeneration ” ” even in the absence of clinical symptoms ” ” and these findings can be evaluated with cartilage - sensitive mri . study design : retrospective cohort . methods : twenty - nine patients (ages : 15 - 40) who had undergone isolated arthroscopic partial medial or lateral meniscectomy with at least 5 - year follow - up were evaluated . all patients had arthroscopically normal articular cartilage at the time of initial meniscectomy . seventeen patients (18 knees) underwent partial medial meniscectomy (mm) and 12 patients underwent partial lateral meniscectomy (lm) with mean follow - up of 8 . 4 and 7 . 1 years ” ” respectively . follow - up evaluation included physical examination ” ” outcome questionnaires ” ” and cartilage - sensitive mri examination with modified outerbridge grading of articular surfaces . results : outerbridge grades ii - iv were noted in 64 % of medial compartment joint surfaces in group mm knees versus 33 % of lateral compartment joint surfaces in group lm knees . abnormal cartilage surfaces (grades ii - iv ”) ” subchondral sclerosis ” ” and condylar squaring were all significantly more frequent after medial meniscectomy ($p < 0 . 05$) . groups mm and lm had no significant differences among outcome scores ” ” which remained excellent in both groups . a significant negative correlation was found between the severity of cartilage wear and functional scoring in the mm group ” ” suggesting that functional disability lags behind early mri evidence of degeneration

0.100137 9192314 1 . we have investigated the roles of adenosine in regulating the respiratory and cardiovascular systems of rats that were made chronically hypoxic for 3 - 4 weeks from 6 weeks of age (ch rats) in an hypoxic chamber at 12 % o₂ . they were studied under anaesthesia while breathing 12 % o₂ and during acute hypoxia (breathing 8 % o₂ for 5 min) before and after addition of the adenosine receptor antagonist 8 - phenyltheophylline (8 - pt " " 10 mg kg⁻¹). the results were compared with those obtained from normoxic (n) rats in a previous study . 2 . ch rats breathing 12 % o₂ had greater minute ventilation (vp) than n rats breathing air " " but their levels of arterial blood pressure (abp ") " heart rate (hr ") " femoral vascular conductance (fvc) and cerebral vascular conductance (cvc) were fully comparable . 8 - pt increased tidal volume (vt) in ch rats indicating a greater tonic central inhibitory influence of adenosine on vt than in n rats . however " " 8 - pt had no effect on cardiovascular variables " " indicating no tonic cardioinhibitory or vasodilator influence of adenosine in ch rats . 3 . acute hypoxia in ch rats increased ve such that at the 5th minute of 8 % o₂ absolute ve was comparable to that of n rats breathing 8 % o₂ . moreover " " in ch rats 8 - pt increased vt at the 5th minute of 8 % o₂ indicating that the central inhibitory influence of adenosine limits the ability to maintain vt in acute hypoxia as it does in n rats . 4 . eight per cent o₂ also produced a full in abp in ch rats that was comparable to that induced in n rats by the larger change from air to 8 % o₂ . however " " the changes in hr were similar in ch and n rats while the increases in fvc and cvc were smaller in ch rats . this suggests that the ability of the secondary effects of hyperventilation and of the

20723251 it is common knowledge that cocaine has been linked to the development of various acute and chronic cardiovascular complications including acute coronary syndromes . we present a young " " male patient " " drug abuser who underwent cabg due to anterolateral myocardial infarction . our presentation is one of the very rare cases reported in literature regarding acute thrombosis of left main coronary artery related to cocaine use " " in a patient with normal coronary arteries " " successfully operated . drug - abusers seem to have increased mortality and morbidity after surgery and high possibility for stent thrombosis after percutaneous coronary interventions " " because of their usually terrible medical compliance and coexistent several problems of general health . there are no specific guidelines about treatment of thrombus formation in coronary arteries " " as a consequence of cocaine use

-0.013722 12562815 adhesion to both abiotic and biotic surfaces by the gram - negative prothescate bacterium *caulobacter crescentus* is mediated by a polar organelle called the " " " " holdfast " " " " which enables the bacterium to form stable monolayer biofilms . the holdfast " " a complex polysaccharide composed in part of n - acetylglucosamine " " localizes to the tip of the stalk (a thin cylindrical extension of the cell wall and membranes). we report here the isolation of adhesion mutants with transposon insertions in an uncharacterized gene cluster involved in holdfast biogenesis (hfs) as well as in previously identified polar development genes (podj and plec ") " and the holdfast attachment genes (hfa). clean deletions of three of the four genes in the hfs gene cluster (hfsdab) resulted in a severe holdfast biogenesis phenotype . these mutants do not bind to surfaces or to a fluorescently labeled lectin " " specific for n - acetylglucosamine . transmission electron microscopy indicated that the hfsdab mutants fail to synthesize a holdfast at the stalk tip . the predicted hfs gene products have significant sequence similarity to proteins necessary for exopolysaccharide export in gram - negative bacteria . hfsa has sequence similarity to gumc from *xanthomonas campestris* " " which is involved in exopolysaccharide export in the periplasm . hfsd has sequence similarity to wza from *escherichia coli* " " an outer membrane protein involved in secretion of polysaccharide through the outer membrane . hfsb is a novel protein involved in holdfast biogenesis . these data suggest that the hfs genes play an important role in holdfast export .

23144810 there is a lack of large - scale studies on vitamin d status and its relationship to parathyroid hormone (pth) and bone turnover markers in adults living in shanghai . the objectives were to determine the prevalence of vitamin d insufficiency in shanghai and to investigate the relationship of 25 (oh) d with parathyroid function and bone turnover markers . this cross - sectional study involved 649 men and 1939 women aged 20 - 89 years who were randomly sampled in shanghai . serum concentrations of 25 (oh) d " " pth " " albumin " " and bone turnover markers were measured . during the winter season " " the prevalence of vitamin d insufficiency (≥ 30 ng / ml) was 84 % in males and 89 % in females . the prevalence of vitamin d deficiency (≥ 20 ng / ml) was 30 % in males and 46 % in females . with increasing serum 25 (oh) d concentrations categorized as ≥ 10 " " 10 - 20 " " 20 - 30 " " and 30 ng / ml " " the mean pth and bone turnover markers levels gradually decreased in both sexes ($p < 0.001$). there was an inverse relationship between the serum 25 (oh) d and pth concentrations in both genders " " but no threshold of 25 (oh) d at which pth levels plateaued was observed . there were modest but significantly inverse relationships between the levels of 25 (oh) d and bone turnover markers " " but no plateau was observed for serum 25 (oh) d levels up to 40 ng / ml .